

38098

**Delaval, Jan**

**Fr m:** Roark, Jessica  
**Sent:** Monday, March 19, 2001 2:51 PM  
**To:** Delaval, Jan  
**Subject:** 09/726899 search

Hi Jan,

Please search, including interference, from 09/726,899

SEQ ID NO:3 as an OLIGO search.

Results on paper and disk please.

Thanks!

Jessica H. Roark  
CM1 9B03  
Art Unit 1644  
703 605-1209



OM protein - protein search, using OM  
 Run on: March 19, 2001, 14:55:27 ; Search time 15.19 Seconds  
 (without alignments)  
 576.642 Million cell updates/sec

US-09-726-899-3

perfect score: 129  
Sequence: 1 MSFPKYKPPSSLRTLPETLDP..

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 195891 seqs, 67900655 residues

size : 0  
satisfying chosen parameters: 195891

Maximum DB seq Length: 0

Maximum DB seq length: 2000000000

post-processing: Listing first 45 summer

```
Database : PIR_66:*
1: pir1:*
```

4: p1r4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysts of the total score distribution.

CINEMARTS

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	129	100.0	129	2	JF0383	NADH dehydrogenase
2	17	13.2	129	2	S58237	NADH ubiquinone ox
3	8	6.2	133	2	UC2003	hypothetical prote
4	8	6.2	160	2	B71176	hypothetical prote
5	8	6.2	179	2	A33164	glutamate--tRNA 11
6	8	6.2	368	2	S38609	glutamate--tRNA 11
7	8	6.2	451	2	H82044	C-4-diboxylate t
8	8	6.2	468	2	E83046	L-seryl-tRNAK(ser)
9	8	6.2	658	2	T40107	multifunctional am
10	8	6.2	1440	1	SYHUQT	hypothetical prote
11	7	5.4	227	2	T22144	34.5K linker polyp
12	7	5.4	278	2	C41841	prephenate dehydra
13	7	5.4	310	2	T36068	luciferase XF1379
14	7	5.4	321	2	G82688	hypothetical prote
15	7	5.4	400	2	T71205	hypothetical prote
16	7	5.4	587	2	140849	succinate dehydrog
17	7	5.4	608	2	A64952	sensor protein Ato
18	7	5.4	628	2	S77374	cell division prot
19	7	5.4	748	2	148744	semaphorin A - mou
20	7	5.4	967	2	H71087	leucyl-tRNA 11ga
21	7	5.4	967	2	H75133	leucyl-tRNA synth
22	7	5.4	111	1	T01078	hypothetical prote
23	7	5.4	129	1	GCFLF	glucagon - bigeye
24	6	4.7	29	2	A61135	glucagon I - Europ
25	6	4.7	29	2	C60840	hypothetical prote
26	6	4.7	57	2	C48642	34.5K linker prote
27	6	4.7	58	2	P00129	glucagon precursor
28	6	4.7	63	1	GCIDC	

45	6	4.7	164	2	D42148	glutacoin precursor
44	6	4.7	159	2	B24691	hypothetical prote
44	6	4.7	158	2	B24691	hypothetical prote
42	6	4.7	155	2	T48503	hypothetical prote
41	6	4.7	152	2	T48503	hypothetical prote
40	6	4.7	153	2	A33090	hypothetical prote
39	6	4.7	145	2	E83573	conglutin delta pr
38	6	4.7	133	2	T36401	hypothetical prote
37	6	4.7	132	2	S18465	hypothetical prote
36	6	4.7	122	1	S78256	matrix protein M2
35	6	4.7	108	2	GCAE2	ribosomal protein
34	6	4.7	99	2	T22320	glucanase 2 precurs
33	6	4.7	97	2	A69514	hypothetical prote
32	6	4.7	93	2	S76406	hypothetical prote
31	6	4.7	87	1	D69960	hypothetical prote
30	6	4.7	80	2	B23617	glucanase precursor

## ALIGNMENTS

RESULT 1  
JE0383  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain NDUPB4 human  
NADH dehydrogenase (ubiquinone) oxidoreductase subunit NDUPB4  
N:Alternate names: NADH:ubiquinone oxidoreductase subunit NDUPB4  
C:Species: Homo sapiens (man)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: JE0383  
R:Loeffen, J.L.C.M.; Triefels, R.H.; van den Heuvel, L.P.; Schuelke, M.; Buskens, C.A.  
Biochem. Biophys. Res. Commun. 253, 415-422, 1998  
A:Title: cDNA of eight nuclear subunits of NADH:ubiquinone oxidoreductase: Hu  
A:Reference number: JE0379; MUID:99097250  
A:Accession: JE0383  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-129 <UOE>  
A:Cross-references: GB:AF044957; NID:94164445; PIDN:AA05421.1; PID:94164446  
C:Keywords: NAD; oxidoreductase

DB 2: Length 129;

Query Match	Score 125	Indels 0	Gaps 0
	100.0%;		
	100.0%;		
	100.0%;		
	Pred. No. 1.7e-126;		

Query: ...  
 Best Local Similarity 100.00; Mismatches 0;  
 Matches 129; Conservative

1 MSFPPKSSRLTLPETLPAEYNISPETRAOERLAIRAO LNKEL LDK 60

MSSEPKYKPSLRTLPETLDPAEYNISPETRAQAERLAIRQLKREILEELVLR  
100

MSSEPKYKPSLRTLPETLDPAEYNISPETRAQAERLAIRQLKREILEELVLR  
120

Db  
1 MSLEIN-----  
21 ENDAIIIRWAVARTINVPNFRPTPKNSLMGALCGFPLFIYYIIKENDRNAND-  
120

01 ENT FILLER.....  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 00  
 01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 00  
 01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 00  
 01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 00  
 01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 00  
 01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84

**D6**

61 ENFALIN.....  
SY 129

121	LDRIFFUS	22
111111111		
129		

Db 121 LDRTHLSI 12,

RESULT 2 . . . n15 - bovine

ubiquinone (EC 1.6.5.3) chain 10  
528237  
05-NOV-19

C;Species: Bos primigenius taurus \text\\_cunning-  
13-Jan-1995 #sequence\\_revision 13-Jan-1995  
13-Jan-1995 #text\\_cunning-

C>Date: 1980  
C/Date: 1980  
C/Accession: S28237  
C/Author: T.F.; Arizmendi, J.M.; Dupuis, A.; Fearnley, L.M.; ...  
C/Title: ... phosphatase from ...

R. Walker, J. E. 1992, 1051-1072, 1992  
J. Mol. Biol. 226, 1051-1072, 1992  
sequences of 20 subunits of NADH: ubiquinone oxidoreductase

A; Title: Sequences  
A; Reference number: S28237; PMID:9238931/  
00037

A: Accession: 526237  
A: Status: preliminary

A:Molecule type: unknown  
A:Residues: 1-129 <WALL>  
name: Y6A898: NID:g113; PIDN:CAA6107.1; PID:g114  
func: oxidoreductase

**Keywords:** electron transfer; mitochondrion; NAD; cross-references: EMBL; NCBI; PubMed; A; Cross-references: EMBL; NCBI; PubMed; A; Keywords: electron transfer; mitochondrion; NAD;

100

Query Match

Best Local Similarity 13.2%; Score 17; DB 2; Length 129;  
Matches 17; Conservative 100.0%; Pred. No. 3.7e-10;

Oy 109 RDRKELLOEKLDRTF 125  
Db 109 RDRKELLOEKLDRTF 125

RESULT 3

JC2003

NADH ubiquinone oxidoreductase B15 chain like protein - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 14-Jul-1994 #sequence,revision 14-Jul-1994 #text\_change 07-Feb-1997  
C:Accession: J02003  
R:Goldberg, G.S.; Kaczmarczyk, W.  
A:Reference number: JC2003; MUID:94040816  
A:Accession: JC2003  
A:Molecule type: DNA  
A:Residues: 1-133 <COL>  
C:Genetics:  
A:Gene: 99hpw  
A:Introns: 67/3  
C:Keywords: homeobox; phosphoprotein; sulfoprotein; transmembrane protein  
F:95-112/Domain: transmembrane #status predicted <TM>  
F:30/Binding site: sulfate (Tyr) (covalent) #status predicted  
F:114/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

Query Match

Best Local Similarity 6.2%; Score 8; DB 2; Length 133;  
Matches 8; Conservative 100.0%; Pred. No. 0.85;

Oy 34 AERLAIRA 41  
Db 41 AERLAIRA 48

RESULT 4

B71176

hypothetical protein PH1689 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence,revision 14-Aug-1998 #text\_change 21-Jul-2000  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5: 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Accession: B71176  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-160 <KAW>  
A:Cross-references: GB:AF000007; NID:93236134; PIDN:BA30801.1; PID:93258118  
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1689  
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1689

Query Match

Best Local Similarity 6.2%; Score 8; DB 2; Length 160;  
Matches 8; Conservative 100.0%; Pred. No. 1;

Oy 7 KPSSLRTL 14  
Db 75 KPSSLRTL 82

RESULT 5

A33164

hypothetical protein water - chicken  
N:Alternate names: homeotic protein Hox-3.1 homolog; hypothetical 13.1k protein  
C:Species: Gallus gallus (chicken)  
C:Date: 28-Mar-1991 #sequence,revision 28-Mar-1991 #text\_change 21-Jul-2000  
C:Accession: A33164; JC1395  
R:Goldberg, G.S.  
A:Reference number: A33164  
A:Accession: A33164  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-179 <GO2>  
A:Experimental source: strain Leghorn  
R:Goldberg, G.S.; Kaczmarczyk, W.  
A:Title: Sequence of a novel chicken genomic DNA fragment that hybridizes to the murine

Query Match

Best Local Similarity 6.2%; Score 8; DB 2; Length 179;  
Matches 8; Conservative 100.0%; Pred. No. 1.1;

Oy 34 AERLAIRA 41  
Db 55 AERLAIRA 62

RESULT 6

S38809

glutamate--tRNA ligase (EC 6.1.1.17) - human (fragment)  
N:Alternate names: glutamyl-tRNA synthetase  
C:Species: Homo sapiens (man)  
C:Date: 29-Jul-1997 #sequence,revision 29-Aug-1997 #text\_change 10-Jul-1998  
R:Kaiser, E.; Eberhard, D.; Knippers, R.  
J. Mol. Evol. 34: 45-53, 1992  
A:Title: Exons encoding the highly conserved part of human glutamyl-tRNA synthetase  
A:Accession: S38809; MUID:92211721  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
C:Superfamily: human multifunctional amino acid--tRNA ligase; amino acid--tRNA ligase  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis  
F:31-307/Domain: glutamine--tRNA ligase homology <EG2>

Query Match

Best Local Similarity 6.2%; Score 8; DB 2; Length 368;  
Matches 8; Conservative 100.0%; Pred. No. 2;

Oy 113 EKLIOECK 120  
Db 114 EKLIOECK 121

RESULT 7

H82044

C4-dicarboxylate transporter, anaerobic VC2699 [imported] - Vibrio cholerae  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence,revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: H82044  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82044  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-451 <HEI>  
A:Cross-references: GB:AE004335; GB:AE003852; NID:g9657289; PIDN:AAF95840.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2699  
A:Map position: 1  
C:Superfamily: dicarboxylate membrane-transporter protein A

Query Match 6.2%; Score 8; DB 2; Length 451;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GLENPAL 65  
DB 267 GLENPAL 274

## RESULT 8

E83046  
L:serY1-CRNA(ser) selenium transferase PA4808 [imported] - *Pseudomonas aeruginosa* (strain  
C:Species: *Pseudomonas aeruginosa*  
C>Date: 15-Sep-2000 #sequence, revision 15-Sep-2000 #text\_change 15-Sep-2000  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
..; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A:Reference number: A82950  
A:Accession: E83046  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <STO>  
A:Cross-references: GB:AE004893; GB:AE004091; NID:g9951063; PIDN:AA08194.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: selA; PA4808

Query Match 6.2%; Score 8; DB 2; Length 468;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 RAQAEFLA 38  
DB 361 RAQAEFLA 368

## RESULT 9

TA0107  
Hypothetical 57.9 kd protein - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C>Date: 03-Dec-1999 #sequence, revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: TA0107  
R:Wood, V.; Rajandream, M.A.; Barrall, B.G.; Taylor, K.; Harris, D.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21906  
A:Accession: TA0107  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-658 <WOO>  
A:Cross-references: EMBL:AL031788; PIDN:CAA21162.1; GSPDB:GN00067; SPDB:SPBC2D10.04  
A:Experimental source: strain 97Zh-; cosmid c2D10  
C:Genetics:  
A:Gene: SPDB:SPBC2D10.04

A:Map position: 2

Query Match 6.2%; Score 8; DB 2; Length 658;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 IENPALR 67  
DB 170 IENPALR 177

## RESULT 10

SYHUQT  
multifunctional aminoacyl-tRNA ligase - human  
N:Alternate names: glutroyl-tRNA synthase; multifunctional aminoacyl-tRNA synthetase  
N:Contents: glutamate-tRNA ligase (EC 6.1.1.17); proline-tRNA ligase (EC 6.1.1.15)  
C:Species: *Homo sapiens* (man)  
C>Date: 30-Jun-1992 #sequence, revision 30-Jun-1992 #text\_change 18-Jun-1999  
C:Accession: A38663; S03424; S00969  
R:Felt, R.; Knippers, R.  
J. Biol. Chem. 266, 1448-1455, 1991  
A:Title: The primary structure of human glutamyl-tRNA synthetase. A highly conserve  
A:Reference number: A38663; MUID:91107633  
A:Accession: A38663  
A:Molecule type: mRNA  
A:Residues: 1-1440 <FET>  
A:Cross-references: GB:X54326; NID:g31957; PIDN:CAA38224.1; PID:g31958  
A:Note: the cited Genbank accession number, X54327, is not in release 101.0  
R:Knippers, R.  
submitted to the EMBL Data Library, April 1988  
A:Reference number: S03424  
A:Accession: S03424

A:Molecule type: mRNA  
A:Residues: 96-382, 'U', 384-497, 'IGATSTLO', 506, 'YT', 509, 'WQME', 514, 'SYL', 518, 'WQSLWKT  
A:Cross-references: EMBL:X07466; NID:g31769; PIDN:CAA30334.1; PID:g825664  
R:Thommes, P.; Felt, R.; Schray, B.; Kunze, N.; Knippers, R.  
Nucleic Acids Res. 16, 5391-5406, 1988  
A:Title: The core region of human glutamyl-tRNA synthetase homologues with the Esch  
A:Reference number: S00969; MUID:88262551  
A:Accession: S00969  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 96-146, 'I', 148-191, 'T', 193-382, 'U', 384-416, 'G', 418-497, 'IGATSTLO', 506, 'YT  
HO>

A:Cross-references: EMBL:X07466  
R:Cerini, G.; Kerjan, P.; Astier, M.; Gratecos, D.; Mirande, M.; Semeriva, M.  
EMBO J. 10, 4267-4277, 1991  
A:Title: A component of the multisynthetase complex is a multifunctional aminoacyl-tr  
A:Reference number: S18644; MUID:92097547  
A:Contents: annotation; demonstration of glutamyl- and prolyl- tRNA synthase activiti  
C:Genetics:  
A:Gene: GDB:EPBS; QPRS; QARS  
A:Cross-references: GDB:126609; OMIM:138295  
A:Map position: 1932-1942  
C:Superfamily: human multifunctional amino acid-tRNA ligase; amino acid-tRNA ligase  
C:Keywords: aminoacyl-tRNA synthetase; ATP; duplication; ligase; protein biosynthesis  
F:125-485/Domain: tRNA-charging <GLN>  
F:126-402/Domain: glutamine-tRNA ligase homology <EGL>  
F:677-733/Region: 57-residue repeat  
F:688-733/Domain: amino acid-tRNA ligase repeat homology <ATL1>  
F:750-806/Region: 57-residue repeat  
F:761-806/Domain: amino acid-tRNA ligase repeat homology <ATL2>  
F:828-884/Region: 57-residue repeat  
F:839-884/Domain: amino acid-tRNA ligase repeat homology <ATL3>

Query Match 6.2%; Score 8; DB 1; Length 1440;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 EKLIOEGR 120  
IIIIIIIIII

Db 209 EKLIOEGK 216

## RESULT 11

T22144  
hypothetical protein F43G9.5 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22144

R:Kershaw, J.

submitted to the EMBL Data Library, September 1996

A:Reference number: Z19523

A:Accession: T22144

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-227 &lt;NTL&gt;

A:Cross-references: EMBL:Z79755; PIDN:CAB02106.1; GSPDB:GN00019; CESP:F43G9.5

A:Experimental source: clone F43G9

C:Accession: T22144

A:Residues: 1-227 &lt;NTL&gt;

A:Introns: 70/3; 103/2; 180/1

## Query Match

Best Local Similarity 5.4%; Score 7; DB 2; Length 227;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 RTINVP 78

DB 32 RTINVP 38

## RESULT 12

C41841

34.5K linker polypeptide, phycoerythrocyanin-associated, rod - *Anabaena* sp. (strain PCCC:Species: *Anabaena* sp.

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999

C:Accession: C41841

R:Swanson, R.V.; de Lorimier, R.; Glazer, A.N.

J. Bacteriol. 174, 2640-2647, 1992

A:Title: Genes encoding the phycoobilisome rod substructure are clustered on the *Anabaena*

A:Reference number: A1841; MUID:92210509

A:Accession: C41841

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-278 &lt;SMA&gt;

A:Cross-references: GB:M80357; NID:9142069; PIDN:AAA22018.1; PID:9142072

C:Family: phycoerythrin linker protein cpch3

## Query Match

Best Local Similarity 5.4%; Score 7; DB 2; Length 278;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 AERLAIR 40

DB 6 AERLAIR 12

## RESULT 13

T36068

prephenate dehydratase - *Streptomyces coelicolor*C:Species: *Streptomyces coelicolor*

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000

C:Accession: T36068

R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Burrell, B.G.; Rajandream, M.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z21595

A:Accession: T36068

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-310 &lt;SAU&gt;

A:Cross-references: EMBL:AL034355; PIDN:CAA22234.1; GSPDB:GN00070; SCORDB:SCD78.29C  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: pheA; SCORDB:SCD78.29C  
C:Superfamily: prephenate dehydratase; prephenate dehydratase homology

## Query Match

Best Local Similarity 5.4%; Score 7; DB 2; Length 310;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 LRTLPET 17

DB 19 LRTLPET 25

## RESULT 14

G82688

luciferase Xf1379 (imported) - *Xylella fastidiosa* (strain 9a5c)C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: G82688

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: G82688

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-321 &lt;SIM&gt;

A:Cross-references: GB:AE003969; GB:AE003849; NID:9106379; PIDN:AAF84188.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Birones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neiro, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.

Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: Xf1379

## Query Match

Best Local Similarity 5.4%; Score 7; DB 2; Length 321;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 AERLAIR 40

DB 226 AERLAIR 232

## RESULT 15

T31205

hypothetical protein 666 - *Sphingomonas aromaticivorans* plasmid pNL1C:Species: *Sphingomonas aromaticivorans*

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000

C:Accession: T31205

R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.

submitted to the EMBL Data Library, July 1998

A:Description: Complete sequence of a 184 kb catabolic plasmid from *Sphingomonas arom*

A:Reference number: Z20992

A:Accession: T31205

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-347 <ROM>  
 A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378346; PIDN:AMD03929.1  
 C:Genetics:  
 A:Genome: plasmid pNL1  
 A:Note: orf666  
 C:Superfamily: Sphingomonas aromaticivorans plasmid pNL1 hypothetical protein 666

Query Match 5.4%; Score 7; DB 2; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 AQAERLA 38  
 |||||  
 Db 328 AQAERLA 334

Search completed: March 19, 2001, 14:56:05  
 Job time: 38 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2001, 14:55:52 ; Search time 9.82 Seconds  
(without alignments)  
424.230 Million cell updates/sec

Title: US-09-726-899-3

Perfect score: 129  
Sequence: 1 MSFPPKPSRLTLPETLDP.....DRKEKIQGKIDRTFHLISY 129

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 88757 seqs, 32294092 residues

Word size: 0

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	99.2	128	1	NBSM_HUMAN
2	128	13.2	128	1	NBSM_BOVIN
3	8	6.2	133	1	NBSM_CHICK
4	8	6.2	1440	1	SYEP_HUMAN
5	7	5.4	277	1	PYR2_ANASP
6	7	5.4	398	1	HISX_SOUSO
7	7	5.4	577	1	DCPY_ASPPA
8	7	5.4	587	1	DHSA_COXBU
9	7	5.4	608	1	ATOS_ECOLI
10	7	5.4	628	1	FTH3_SYNY3
11	7	5.4	748	1	SM3B_MOUSE
12	7	5.4	967	1	SVL_PYRHO
13	6	4.7	29	1	GLUC_PLAFA
14	6	4.7	57	1	YAC2_LEGPN
15	6	4.7	71	1	GLUC_ICTPU
16	6	4.7	80	1	CG2L_LUPAN
17	6	4.7	93	1	YOHV_BACSU
18	6	4.7	96	1	GLUC_MYOSC
19	6	4.7	121	1	GLUC_CARAU
20	6	4.7	122	1	GLUC_LOPAM
21	6	4.7	123	1	RL13_OODST
22	6	4.7	123	1	VMAF_VHSVO
23	6	4.7	161	1	V4LG_RHISN
24	6	4.7	164	1	RLB1_RAT
25	6	4.7	180	1	RK6_PORPU
26	6	4.7	181	1	TNPT_PSEPU
27	6	4.7	186	1	RNF_RICER
28	6	4.7	191	1	YK51_RICER
29	6	4.7	199	1	R13A_YEAST
30	6	4.7	199	1	R13B_YEAST
31	6	4.7	201	1	GT_ECOLI
32	6	4.7	201	1	RETB_MOUSE
33	6	4.7	218	1	CCMB_RHOCA

34	6	4.7	227	1	NODW_BRAJA
35	6	4.7	232	1	TONB_CAMCO
36	6	4.7	234	1	PYRE_METAN
37	6	4.7	248	1	P986_MYCNU
38	6	4.7	250	1	PASC_BOBPU
39	6	4.7	251	1	TPIS_PSESY
40	6	4.7	255	1	AC2A_STRCO
41	6	4.7	255	1	YPE1_RHORU
42	6	4.7	274	1	TRY5_ANOCA
43	6	4.7	276	1	RL2_BACSU
44	6	4.7	278	1	PYR2_MASIA
45	6	4.7	279	1	CYL_RHOCA

## ALIGNMENTS

RESULT	ID	Sequence	Score	Length	DB	Description
1	NBSM_HUMAN	128 AA	99.2%	128	1	NBSM_HUMAN
2	NBSM_HUMAN	133 AA	13.2%	133	1	NBSM_BOVIN
3	NBSM_HUMAN	1440 AA	6.2%	1440	1	SYEP_HUMAN
4	NBSM_HUMAN	277 AA	5.4%	277	1	PYR2_ANASP
5	NBSM_HUMAN	398 AA	5.4%	398	1	HISX_SOUSO
6	NBSM_HUMAN	577 AA	5.4%	577	1	DCPY_ASPPA
7	NBSM_HUMAN	587 AA	5.4%	587	1	DHSA_COXBU
8	NBSM_HUMAN	608 AA	5.4%	608	1	ATOS_ECOLI
9	NBSM_HUMAN	628 AA	5.4%	628	1	FTH3_SYNY3
10	NBSM_HUMAN	748 AA	5.4%	748	1	SM3B_MOUSE
11	NBSM_HUMAN	967 AA	5.4%	967	1	SVL_PYRHO
12	NBSM_HUMAN	29 AA	4.7%	29	1	GLUC_PLAFA
13	NBSM_HUMAN	57 AA	4.7%	57	1	YAC2_LEGPN
14	NBSM_HUMAN	71 AA	4.7%	71	1	GLUC_ICTPU
15	NBSM_HUMAN	80 AA	4.7%	80	1	CG2L_LUPAN
16	NBSM_HUMAN	93 AA	4.7%	93	1	YOHV_BACSU
17	NBSM_HUMAN	96 AA	4.7%	96	1	GLUC_MYOSC
18	NBSM_HUMAN	121 AA	4.7%	121	1	GLUC_CARAU
19	NBSM_HUMAN	122 AA	4.7%	122	1	GLUC_LOPAM
20	NBSM_HUMAN	123 AA	4.7%	123	1	RL13_OODST
21	NBSM_HUMAN	161 AA	4.7%	161	1	VMAF_VHSVO
22	NBSM_HUMAN	164 AA	4.7%	164	1	V4LG_RHISN
23	NBSM_HUMAN	180 AA	4.7%	180	1	RLB1_RAT
24	NBSM_HUMAN	181 AA	4.7%	181	1	TNPT_PSEPU
25	NBSM_HUMAN	186 AA	4.7%	186	1	RNF_RICER
26	NBSM_HUMAN	191 AA	4.7%	191	1	YK51_RICER
27	NBSM_HUMAN	199 AA	4.7%	199	1	R13A_YEAST
28	NBSM_HUMAN	199 AA	4.7%	199	1	R13B_YEAST
29	NBSM_HUMAN	201 AA	4.7%	201	1	GT_ECOLI
30	NBSM_HUMAN	201 AA	4.7%	201	1	RETB_MOUSE
31	NBSM_HUMAN	218 AA	4.7%	218	1	CCMB_RHOCA

Query Match: 99.2%, Score 128; DB 1; Length 128;  
Best Local Similarity: 100.0%; Pred. No. 3.4e-125;  
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	SFPKPKPSRLTLPETLDPAEYNISPEPTRRAAEALRAQAKREYLLOYNPNRGLIE	61
DB	1	SFPKPKPSRLTLPETLDPAEYNISPEPTRRAAEALRAQAKREYLLOYNPNRGLIE	60
OY	62	NPALLRMVYARTINYPNRPKNSLMGALCGFPLFIYIITERRDKREKLEOEGKL	121
DB	61	NPALLRMVYARTINYPNRPKNSLMGALCGFPLFIYIITERRDKREKLEOEGKL	120

```

OY 122 DRTFLSY 129
    |||||
DB 121 DRTFLSY 128

RESULT 2
NB5M_BOVIN STANDARD; PRT: 128 AA.
ID NB5M_BOVIN
AC P48305;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE B15 SUBUNIT (EC 1.6.5.3) (EC 1.6.99.3)
DE (COMPLEX I-B15) (CI-B15).
GN NDUFBA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN 1
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=HEART;
RX MEDLINE=92389317; PubMed=1518044;
RA Walker J.E., Arizumi J.M., Dupuis A., Fearley I.M., Finel M.,
RA Medd S.M., Pilkington S.J., Runswick M.J., Skehel J.M.;
RT "Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from
RT bovine heart mitochondria. Application of a novel strategy for
RT sequencing proteins using the polymerase chain reaction."
RL J. Mol. Biol. 226:1051-1072(1992).
CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; MATRIX SIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64898; CAA46107.1;
KW Oxidoreductase; NAD; ubiquinone; Mitochondrion; Acetylation.
FT INIT_MET 0
FT ND_RES 1 ACETYLATION;
FT SEQUENCE 128 AA; 15053 MW; CCI352E9E80DF7D5 CRC64;

Query Match 13.2%; Score 17; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 2e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 109 RDRKEKLIQEGKLDRTF 125
    |||||
DB 108 RDRKEKLIQEGKLDRTF 124

RESULT 3
NB5M_CHICK STANDARD; PRT: 133 AA.
ID NB5M_CHICK
AC P48306;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE NADH-UBIQUINONE OXIDOREDUCTASE B15 SUBUNIT (EC 1.6.5.3)
DE (EC 1.6.99.3) (COMPLEX I-B15) (CI-B15) (HYPOTHETICAL PROTEIN WALTER)
DE (GGPFW).
GN NDUFBA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=94040816; PubMed=7901127;
RA Goldberg G.S., Kaczmarezyk W.;
RT "A chicken genomic DNA fragment that hybridizes to the murine Hox-3.1
RT homeobox is likely to encode the NADH ubiquinone oxidoreductase
RT subunit B15."
RL Gene 133:233-235(1993).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=93077061; PubMed=1359990;
RA Goldberg G.S., Kaczmarezyk W.;
RT "Sequence of a novel chicken genomic DNA fragment that hybridizes to
RT the murine Hox-3.1 homeobox."
RL Gene 121:397-398(1992).
CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; MATRIX SIDE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60778; CAA43193.1; ALT_SEQ.
KW Oxidoreductase; NAD; ubiquinone; Mitochondrion.
FT SEQUENCE 133 AA; 14938 MW; ED7E82948C97B352 CRC64;

Query Match 6.2%; Score 8; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 AERLAIRA 41
    |||||
DB 41 AERLAIRA 48

RESULT 4
SYEP_HUMAN STANDARD; PRT: 1440 AA.
ID SYEP_HUMAN
AC P07814;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MULTIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE [INCLUDES: GLUTAMYL-TRNA
DE SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)]; PROLYL-TRNA
DE SYNTHETASE (EC 6.1.1.15) (PROLINE--TRNA LIGASE)].
GN EPRS OR GPRS OR GLNS OR PARS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=91107633; PubMed=1988429;
RA Pett R., Knippers R.;
RT "The primary structure of human glutamyl-tRNA synthetase. A highly
RT conserved core, amino acid repeat regions, and homologies with
RT translation elongation factors."
RL J. Biol. Chem. 266:1448-1455(1991).
RN 12
RP PRELIMINARY SEQUENCE OF 96-887 FROM N.A.
RC TISSUE=CERVIX CARCINOMA;

```

RX MEDLINE-88262551; PubMed-3290852;  
 RA Theommes P., Felt R., Schray B., Kunze N., Knippers R.;  
 RT "The core region of human glutaminyl-tRNA synthetase homologues with  
 the Escherichia coli and yeast enzymes.";  
 RL Nucleic Acids Res. 16:5391-5406(1988)  
 CC  
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +  
 CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU);  
 CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +  
 CC PYROPHOSPHATE + L-PROLYL-TRNA(PRO);  
 CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH CONTAINS  
 CC NINE DIFFERENT AA-TRNA SYNTHETASES (ARG, ASP, GLU, GLN, ILE, LEU,  
 CC LYS, MET AND PRO).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I  
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II  
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 "WHEP-TRS" DOMAINS.  
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A GLUTAMINYL-TRNA  
 CC SYNTHETASE.  
 CC -1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A NUMBER OF  
 CC SEQUENCE ERRORS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X54326; CA38224.1; -; ALT\_SEQ.  
 CC EMBL: X07466; CA30354.1; ALT\_SEQ.  
 CC PIR: A38663; SYHUT.  
 CC HSSP: P00962; 10RU.  
 CC MIM: 138295; -;  
 CC DR INTERPRO: IPR000738; -;  
 CC DR INTERPRO: IPR000924; -;  
 CC DR INTERPRO: IPR001412; -;  
 CC DR INTERPRO: IPR002106; -;  
 CC DR INTERPRO: IPR002314; -;  
 CC DR INTERPRO: IPR002316; -;  
 CC DR PFAM: PF00458; WHEP-TRS; 3.  
 CC DR PFAM: PF00749; tRNA-synC\_1;  
 CC DR PFAM: PF00587; tRNA-synC\_2b; 1.  
 CC DR PRINTS: PR00987; TRNASYNTHGLU.  
 CC DR PRINTS: PR01046; TRNASYNTHPRO.  
 CC DR PROSITE: PS00178; AA-TRNA\_LIGASE\_1; 1.  
 CC DR PROSITE: PS00179; AA-TRNA\_LIGASE\_1; 1.  
 CC DR PROSITE: PS00339; AA-TRNA\_LIGASE\_1; 2; FALSE\_NEG.  
 CC DR PROSITE: PS00762; WHEP-TRS; 3.  
 CC DR PROSITE: PS00762; WHEP-TRS; 3.  
 CC KMW multioacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 CC KMW Multifunctional enzyme; Repeat.  
 CC FT DOMAIN 92 687  
 CC FT DOMAIN 688 884  
 CC FT DOMAIN 935 1440  
 CC FT SIMILAR 132 142  
 CC FT SIMILAR 360 364  
 CC FT BINDING 363 363  
 CC FT BINDING 688 733  
 CC FT DOMAIN 761 806  
 CC FT DOMAIN 839 884  
 CC FT DOMAIN 887 919  
 CC FT SEQUENCE 1440 AA; 163026 MW; CAE185A0AA41C204 CRC64;  
 SQ

Query Match 6.2%; Score 8; DB 1; Length 1440;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 113 EKLIOGK 120  
 11111111  
 Db 209 EKLIOGK 216

RESULT 5  
 PYR2\_ANASP  
 ID PYR2\_ANASP STANDARD; PRT; 277 AA.  
 AC P31329;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE PHYCOBILISOME 34.5 KDA LINKER POLYPEPTIDE, PHYCOERYTHROCYANIN-  
 DE ASSOCIATED, ROD.  
 GN PECC.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92210509; PubMed-1556083;  
 RA Swanson R.V., de Lorimer R., Glaizer A.N.;  
 RT "Genes encoding the phycoobilisome rod substructure are clustered on  
 RT the Anabaena chromosome: characterization of the phycoerythrocyanin  
 RT operon.";  
 RL J. Bacteriol. 174:2640-2647(1992).  
 RL [2]  
 RP SEQUENCE OF 1-24.  
 RX MEDLINE-92077441; PubMed-1743523;  
 RA Bryant D.A., Stirewalt V.L., Glauser M., Frank G., Sidler W.,  
 RA Zuber H.;  
 RT "A small multigene family encodes the rod-core linker polypeptides of  
 RT Anabaena sp. PCC7120 phycoobilisomes.";  
 RL Gene 107:91-99(1991).  
 CC -1- FUNCTION: ROD LINKER PROTEIN, ASSOCIATED WITH PHYCOERYTHROCYANIN.  
 CC LINKER POLYPEPTIDES DETERMINE THE STATE OF AGGREGATION AND THE  
 CC LOCATION OF THE DISC-SHAPED PHYCOBILIPROTEIN UNITS WITHIN THE  
 CC PHYCOBILISOME AND MODULATE THEIR SPECTROSCOPIC PROPERTIES IN ORDER  
 CC TO MEDIATE A DIRECTED AND OPTIMAL ENERGY TRANSFER.  
 CC -1- SUBCELLULAR LOCATION: THIS PROTEIN OCCURS IN THE ROD, IT IS  
 CC ASSOCIATED WITH PHYCOERYTHROCYANIN.  
 CC -1- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M80357; AAA22018.1; -;  
 CC PIR: C41841; C41841.  
 CC DR INTERPRO: IPR001297; -;  
 CC DR INTERPRO: IPR001685; -;  
 CC DR PFAM: PF01383; CpCD; 1.  
 CC DR PFAM: PF00427; PBS-linker.poly; 1.  
 CC KMW phycoobilisome; Photosynthesis.  
 CC FT INT\_MET 0  
 CC FT SEQUENCE 277 AA; 31164 MW; 9456A0D1FD54161A CRC64;  
 SQ

Query Match 5.4%; Score 7; DB 1; Length 277;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 AERLAIR 40  
 11111111  
 Db 5 AERLAIR 11

RESULT 6  
 HISX\_SULSO  
 ID HISX\_SULSO STANDARD; PRT; 398 AA.  
 AC O33775;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HISTIDINOL DEHYDROGENASE (EC 1.1.1.23) (HDH).  
 GN HISD.  
 CC Sulfolobus solfataricus.  
 OS Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 1617 / P2;  
 RX MEDLINE-97352708; PubMed-9209067;  
 RA Charlebois R.L., Sengen C.W., Doolittle W.F., Brown J.R.;  
 RT "Evolutionary analysis of the hisGABFDEHI gene cluster from the  
 archaeon Sulfolobus solfataricus P2."  
 RL J. Bacteriol. 179:4429-4432(1997).  
 CC -1- FUNCTION: THIS PROTEIN IS CONSIDERED AS A BIFUNCTIONAL ENZYME,  
 POSSESSING TWO ACTIVE SITES, ONE AN ALCOHOL DEHYDROGENASE AND  
 THE OTHER AN ALDEHYDE DEHYDROGENASE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: L-HISTIDINOL + 2 NAD(+) = L-HISTIDINE +  
 2 NADH.  
 CC -1- PATHWAY: TENTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC, FUNGAL AND PLANTS HDH.  
 CC this SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U82227; AAB63023.1; -  
 CC INTERPRO: IPR001692; -  
 CC PFAM: PF00815; Histidinol\_dh; 1.  
 CC PRINTS: PR00083; H0LDHDEGNASE.  
 CC PROSITE: PS00611; HISOL-DEHYDROGENASE; FALSE\_NEG.  
 CC HISTIDINE biosynthesis. Multifunctional enzyme; Oxidoreductase; NAD.  
 KW SEQUENCE 398 AA; 43635 MW; B969B35227ED564B CRC64;  
 SO  
 Query Match 5.4%; Score 7; DB 1; Length 398;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 101 IYIIKT 107  
 Db 276 IYIIKT 282  
 RESULT 7  
 DE ASPA  
 ID ASPA STANDARD: PRT; 577 AA.  
 AC P51844;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PYROVATE DECARBOXYLASE (EC 4.1.1.1).  
 GN PDC.  
 OS Aspergillus parasiticus.  
 CC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;  
 CC anamorphic Trichocomaceae; Aspergillus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 163 / NRRL 5862 / SU-1;  
 RX MEDLINE-94237455; PubMed-8181725;  
 RA Sanchez V., Vinas I., Roberts I.N., Jeenes D.J., Watson A.J.,  
 RA Archer D.B.;  
 RT "A pyruvate decarboxylase gene from Aspergillus parasiticus";  
 RL FEMS Microbiol. Lett. 117:207-210(1994).  
 CC -1- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).  
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U00967; AAA20440.1; -  
 CC DR HSSP: P06169; 1YPD.  
 CC DR INTERPRO: IPR000399; -  
 CC PFAM: PF00205; TPP\_enzymes; 2.  
 CC PROSITE: PS00187; TPP\_ENZYMES; FALSE\_NEG.  
 CC KW Lyase; Decarboxylase; Flavo-protein; Thiamine pyrophosphate.  
 CC ACT SITE 52 52  
 CC BY SIMILARITY.  
 CC SEQUENCE 577 AA; 64071 MW; 32146A17EC930156 CRC64;  
 SO  
 Query Match 5.4%; Score 7; DB 1; Length 577;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 SURTLP 16  
 Db 349 SURTLP 355  
 RESULT 8  
 DHSX\_COXBU  
 ID DHSX\_COXBU STANDARD: PRT; 587 AA.  
 AC P51034;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE SUCCINATE DEHYDROGENASE FLAVOPROTEIN SUBUNIT (EC 1.3.99.1).  
 GN SDHA.  
 OS Coxiiella burnetii.  
 CC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;  
 CC Coxiella group; Coxiella.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NINE MITE;  
 RX MEDLINE-95212926; PubMed-7698664;  
 RA Heinzen R.A., Mo Y.-Y., Robertson S.J., Mallavia L.P.;  
 RT "Characterization of the succinate dehydrogenase-encoding gene  
 RT cluster (sdh) from the rickettsia Coxiella burnetii."  
 RL Gene 155:27-34(1995).  
 CC -1- CATALYTIC ACTIVITY: SUCCINATE + ACCEPTOR = FUMARATE + REDUCED  
 ACCEPTOR.  
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
 CC -1- SUBUNIT: PART OF AN ENZYME COMPLEX CONTAINING FOUR SUBUNITS: A  
 FLAVOPROTEIN, AN IRON-SULFUR PROTEIN, CYTOCHROME B-556 AND A  
 HYDROPHOBIC PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L33409; AAA74133.1; -  
 CC DR INTERPRO: IPR000464; -  
 CC DR PFAM: PF00890; FAD binding\_2; 1.  
 CC DR PROSITE: PS00504; FRD\_SDH\_FAD\_BINDING; 1.  
 CC KW Tricarboxylic acid cycle; Flavo-protein; FAD; Oxidoreductase;  
 CC Election transport.  
 CC NP\_BIND 12 26  
 CC FT BINDING 47 47  
 CC FT ACT SITE 244 244  
 CC ACT SITE 260 260  
 CC BY SIMILARITY.  
 CC SEQUENCE 587 AA; 65438 MW; 37B264ECC9803DB6 CRC64;  
 SO

Query Match 5.4%; Score 7; DB 1; Length 587;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 ENPALLR 67  
Db 455 ENPALLR 461

RESULT 9  
ATOS.ECOLI STANDARD; PRT; 608 AA.  
AC 006067;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SENSOR PROTEIN ATOS (EC 2.7.3.-).  
GN ATOS.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-93348226; PubMed-8346225;  
RA Canelaklis E.S., Paterakis A.A., Huang S.-C., Panagiotidis C.A.,  
RA Kyriakidis D.A.;  
RT "Identification, cloning, and nucleotide sequencing of the ornithine  
decarboxylase antizyme gene of Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:7129-7133(1993).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-9751358; PubMed-9097040;  
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,  
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
RA Sakito N., Sempel G., Seki Y., Sivasubraman S., Tagami H.,  
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map.";  
RL DNA Res. 3:379-392(1996).  
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM ATOS/ATOC;  
CC MAY ACTIVATE ATOC BY PHOSPHORYLATION.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
CC KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L13078; AAA23449.1; -  
DR EMBL: AE000311; AAC75279.1; -  
DR EMBL: D90851; BAA16015.1; -  
DR ECGENE: EGI1667; ATOS.  
DR INTERPRO: IPR000014; -

DR INTERPRO: IPR000410; -  
DR INTERPRO: IPR000658; -  
DR PFAM: PF00672; DUF5.1.  
DR PFAM: PF00989; PAS.1.  
DR PFAM: PF00512; signal.1.  
DR PRINTS: PR00344; BCTRLENSOR.  
KW Sensory transduction; transferase; kinase; phosphorylation;  
KW Transmembrane; Inner membrane.  
FT TRANSMEM 16 36 POTENTIAL.  
FT TRANSMEM 190 210 POTENTIAL.  
FT MOD\_RES 398 398 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT TRANSMEM 584 604 POTENTIAL.  
SQ SEQUENCE 608 AA; 67789 MW; 4669868f98965c9c CRC64;

OY 32 AQAERLA 38  
Db 381 AQAERLA 387

Query Match 5.4%; Score 7; DB 1; Length 608;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10  
FTTH3\_SYNY3 STANDARD; PRT; 628 AA.  
AC P73437;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELL DIVISION PROTEIN FTSH HOMOLOG 3 (EC 3.4.24.-).  
GN SL1463.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [11]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97061201; PubMed-8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naro K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE  
CC (BY SIMILARITY).  
CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41 (ZINC  
CC METALLOPROTEASE).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D90906; BAA17477.1; -  
DR INTERPRO: IPR000130; -  
DR INTERPRO: IPR000642; -  
DR INTERPRO: IPR001939; -  
DR PFAM: PF00004; AAA.1.  
DR PFAM: PF01434; Peptidase\_M41.1.  
DR PROSITE: PS00142; ZINC\_PROTEASE.1.  
DR PROSITE: PS00674; AAA.1.  
KW Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;  
KW Zinc; Multigene family.

```

FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 21 31 POTENTIAL.
FT DOMAIN 32 119 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 120 142 POTENTIAL.
FT DOMAIN 143 628 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 214 221 ATP (POTENTIAL).
FT METAL 438 438 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 439 439 BY SIMILARITY.
FT METAL 442 442 ZINC (CATALYTIC) (BY SIMILARITY).
SO SEQUENCE 628 AA: 68199 MW: 882563000735BA CRC64;

Query Match 5.4%; Score 7; DB 1; Length 628;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PETLPPA 21
Db 324 PETLPPA 330

RESULT 11
SM3B_MOUSE STANDARD; PRT; 748 AA.
AC 062177;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 3B PRECURSOR (SEMAPHORIN A) (SEMA A).
GN SEMA3B OR SEMA OR SEMA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MRI; TISSUE-EMBRYO;
RX MEDLINE-95267431; PubMed-7748561;
RA Puschel A.W., Adams R.H., Betz H.;
RT "Murine semaphorin D/collapsin is a member of a diverse gene family
RL Neuron 14:941-948(1995).
CC -1- FUNCTION: INHIBITS AXONAL EXTENSION BY PROVIDING LOCAL SIGNALS TO
SPECIFY TERRITORIES INACCESSIBLE FOR GROWING AXONS.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM DAY 10 IN THE EMBRYO. LOW
LEVELS FOUND BETWEEN DAYS 10-12. EXPRESSION PEAKS ON DAY 13 WITH
MODERATE LEVELS FROM THEN UNTIL BIRTH.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X85990, CA55982.1; -
CC INTERPRO: IPR001627; -
CC INTERPRO: IPR003006; -
CC PFM: PF01403; Sema; 1.
CC PFM: PF00047; Ig; 1.
CC MGD: MGI:107561; Sema3b
DR Signal, immunoglobulin domain; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 748 SEMAPHORIN 3B.
FT DOMAIN 239 536 SEMA.
FT DOMAIN 636 716 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 698 702 POLY-GLY.
FT DOMAIN 723 743 ARG-RICH (BASIC).
FT DISULFID 643 709 BY SIMILARITY.

```

```

FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 748 AA: 82894 MW: 1866B7D2397C9305 CRC64;

Query Match 5.4%; Score 7; DB 1; Length 748;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 AQERLA 38.
Db 665 AQERLA 671

RESULT 12
SYL_PYRHO STANDARD; PRT; 967 AA.
AC 058698;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE--TRNA LIGASE) (LEURS).
GN LEUS OR PH0965.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE-98344137; PubMed-9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RL thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
CC DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +
PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AP000004; BAA30062.1; -
CC INTERPRO: IPR002300; -
CC PFM: PF00133; tRNA-synL; 1.
CC PROSITE: PS00178; AA-TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SIMILAR 43 53 "HIGH" REGION.
FT SIMILAR 650 654 "RMSK" REGION.
FT BINDING 653 653 ATP (BY SIMILARITY).
SO SEQUENCE 967 AA: 113949 MW: DEFBD287679E9498 CRC64;

Query Match 5.4%; Score 7; DB 1; Length 967;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 PETRRAO 33
Db 489 PETRRAO 495

RESULT 13
GLUC_PLAFE

```



Query Match 4.7%; Score 6; DB 1; Length 71;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 ETTRAO 33  
      |||||  
Db 15 ETTRAO 20

Search completed: March 19, 2001, 14:57:50  
Job time: 118 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2001, 14:55:28 ; Search time 23.36 seconds  
(without alignments)  
647.253 Million cell updates/sec

Title: US-09-726-899-3  
Perfect score: 129  
Sequence: 1 MSFPPKYSRLRFLPETLDP.....DRKEKLIQEGKIDRTFHLST 129

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Woe: 0  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: SPTREMBL\_15:\*
- 2: sp\_archaea:\*
- 3: sp\_bacteria:\*
- 4: sp\_fungi:\*
- 5: sp\_human:\*
- 6: sp\_invertebrate:\*
- 7: sp\_mhc:\*
- 8: sp\_mammal:\*
- 9: sp\_organelle:\*
- 10: sp\_phage:\*
- 11: sp\_plant:\*
- 12: sp\_rodent:\*
- 13: sp\_virus:\*
- 14: sp\_vertebrate:\*
- 15: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	6.2	160	1	059313 pyrococcus
2	8	6.2	451	2	09KNN2
3	8	6.2	3	3	074798 schizosacch
4	7	5.4	120	5	077180
5	7	5.4	129	11	P70632
6	7	5.4	194	2	09RJI5
7	7	5.4	227	5	093716
8	7	5.4	310	2	092BX0
9	7	5.4	321	2	09PDK0
10	7	5.4	341	10	09LWK8
11	7	5.4	347	2	085914
12	7	5.4	400	5	023568
13	7	5.4	627	4	060826
14	7	5.4	627	11	09JIG7
15	7	5.4	636	3	09Y8A4
16	7	5.4	678	10	09LHNS
17	7	5.4	851	5	09YUVA
18	7	5.4	857	10	09LNM4
19	7	5.4	875	5	09V6A2

20	7	5.4	967	1	09V0B9	09V0B9 pyrococcus
21	7	5.4	1111	10	092OX8	092OX8 arabidopsis
22	7	5.4	2271	5	09W1A9	09W1A9 drosophila
23	7	5.4	4150	2	09KIV4	09KIV4 streptomyce
24	6	4.7	11	6	09TOS9	09TOS9 equus caball
25	6	4.7	20	2	09R636	09R636 desulfovibr
26	6	4.7	79	5	09VY74	09VY74 drosophila
27	6	4.7	81	5	09V655	09V655 drosophila
28	6	4.7	92	2	09KX34	09KX34 pseudomonas
29	6	4.7	97	2	P74434	P74434 synechocyst
30	6	4.7	99	1	028167	028167 archaeoglob
31	6	4.7	100	2	069234	069234 bacillus ce
32	6	4.7	102	10	09LH25	09LH25 oryza sativ
33	6	4.7	108	5	020476	020476 caenorhabdl
34	6	4.7	109	12	081628	081628 hepatitis c
35	6	4.7	109	12	081629	081629 hepatitis c
36	6	4.7	117	5	09XZ86	09XZ86 drosophila
37	6	4.7	119	8	09MEK5	09MEK5 leplemur s
38	6	4.7	121	2	P71515	P71515 methylodact
39	6	4.7	125	2	09K417	09K417 streptomyce
40	6	4.7	125	10	09SU24	09SU24 arabidopsis
41	6	4.7	133	1	093751	093751 methanother
42	6	4.7	133	2	09RK23	09RK23 streptomyce
43	6	4.7	135	12	065324	065324 bovine aden
44	6	4.7	138	5	09VW05	09VW05 drosophila
45	6	4.7	140	5	09N923	09N923 drosophila

## ALIGNMENTS

RESULT 1  
ID 059313 PRELIMINARY; PRT: 160 AA.  
AC 059313;  
DT 01-AUG-1998 (TEMBLrel. 07, Created)  
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 17.4 KDA PROTEIN PH1689.  
GN PH1689.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
ON NCBI\_TaxID=53953;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida M., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.,  
RT Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL: AP000007; BAA30801.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 160 AA; 17448 MW; 5BC6EDB83B244A9A CRC64;

Query Match 6.2%; Score 8; DB 1; Length 160;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 KPSSLRTL 14  
DB 75 KPSSLRTL 82  
RESULT 2  
ID 09KNN2 PRELIMINARY; PRT: 451 AA.

AC 09KNN2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE C4-DICARBOXYLATE TRANSPORTER, ANAEROBIC.  
 GN VC2699.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=566;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N1661 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae."  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004335; AAF95840.1; -.  
 DR TIGR; VC2699; -.  
 SQ SEQUENCE 451 AA; 47118 MW; CA867A96A507AC60 CRC64;

Query Match 6.2%; Score 8; DB 2; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 3.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 GLIENPAL 65  
 DB 267 GLIENPAL 274

RESULT 3  
 074798 PRELIMINARY; PRT: 658 AA.  
 AC 074798;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHEICAL 57.9 KDA PROTEIN CD210.04 IN CHROMOSOME II.  
 GN SPC2D10.04.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
 OC Schizosaccharomycetaceae; Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: TO YEAST YJL084L.  
 DR EMBL; AL031788; CAA21162.1; -.  
 SQ SEQUENCE 658 AA; 72731 MW; 40EB06BB970B4F7 CRC64;

Query Match 6.2%; Score 8; DB 3; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 IENPALR 67  
 DB 170 IENPALR 177

RESULT 4  
 077180 PRELIMINARY; PRT: 120 AA.  
 AC 077180;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE ERYTHROCYTE MEMBRANE PROTEIN 1 TYPE Q (FRAGMENT).  
 GN VAR.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1776;  
 RA Tiwari B.;  
 RT "Plasmodium falciparum parasites with short forms of chromosome 9  
 express var genes and PfEMP1 protein."  
 RL submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF084588; AAC62730.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 120 AA; 14437 MW; 8330E18A2C09596C CRC64;

DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE ERYTHROCYTE MEMBRANE PROTEIN 1 TYPE Q (FRAGMENT).  
 GN VAR.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1776;  
 RA Tiwari B.;  
 RT "Plasmodium falciparum parasites with short forms of chromosome 9  
 express var genes and PfEMP1 protein."  
 RL submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF084588; AAC62730.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 120 AA; 14437 MW; 8330E18A2C09596C CRC64;

Query Match 5.4%; Score 7; DB 5; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 ERDRK 114  
 DB 14 ERDRK 20

RESULT 5  
 P70632 PRELIMINARY; PRT: 129 AA.  
 AC P70632;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last annotation update)  
 DE FURROSEMI-SENSITIVE K-CL COTRANSPORTER (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WISTAR-KYOTO;  
 RA Adams L.A., Werny I., Schwartz S.M.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U75396; AAB18960.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 129 AA; 15495 MW; EEA843A9C3B1990E CRC64;

Query Match 5.4%; Score 7; DB 11; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 KTERDRK 112  
 DB 102 KTERDRK 108

RESULT 6  
 09RL5 PRELIMINARY; PRT: 194 AA.  
 AC 09RL5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE PUTATIVE TETR FAMILY TRANSCRIPTIONAL REGULATOR.  
 GN SCF6.16.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Kinsachi H., Hopwood D.A.;  
 RL "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RT M1. Microbiol. 21:77-96(1996).  
 DR EMBL: AL121849; CAB58280.1; -  
 DR INTERPRO: IPR001647; -  
 DR PFAM: PF00440; tetr. 1.  
 DR PRINTS: PRO0455; HTHMTR.  
 SQ SEQUENCE 194 AA; 20842 MW; 8335089D11453DF4 CRC64;

Query Match 5.4%; Score 7; DB 2; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 RAOAERL 37  
 Db 172 RAOAERL 178  
 |||||

RESULT 7  
 ID 093716 PRELIMINARY; PRT; 227 AA.  
 AC 093716;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE F43G9.5 PROTEIN.  
 GN F43G9.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Poloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Parsons J., Percy C., Rifken L., Mortimore B., O'Callaghan M.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."  
 RT Nature 368:32-38(1994).  
 RL EMBL: 279755; CAB02106.1; -  
 DR INTERPRO: IPR001993; -  
 DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
 SQ SEQUENCE 227 AA; 25916 MW; 035A804265BCED48 CRC64;

Query Match 5.4%; Score 7; DB 5; Length 227;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 RTINYP 78  
 Db 32 RTINYP 38  
 |||||

RESULT 8  
 ID 092BX0 PRELIMINARY; PRT; 310 AA.  
 AC 092BX0;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE PREPHENATE DEHYDRATASE.  
 GN SCD78.29C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RC MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinsachi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL M1. Microbiol. 21:77-96(1996).  
 DR EMBL: AL034355; CAA22234.1; -  
 DR INTERPRO: IPR001086; -  
 DR PFAM: PF00800; PDT. 1.  
 DR PROSITE: PS00857; PREPHENATE\_DEHYDR. 1; 1.  
 SQ SEQUENCE 310 AA; 33440 MW; 60D061D53960EBD4 CRC64;

-Query Match 5.4%; Score 7; DB 2; Length 310;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 LRTLPET 17  
 Db 19 LRTLPET 25  
 |||||

RESULT 9  
 ID 09PDK0 PRELIMINARY; PRT; 321 AA.  
 AC 09PDK0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE LUCIFERASE.  
 GN XF1379.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN-9ASC;  
 RX MEDLINE-20365717; PubMed-10910347;  
 RA Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carrero D.M., Carrer H.,  
 RA Coutinho N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,  
 RA Facincani L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Fraga J.S., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Gantier M., Goldman S.C., Franco M.C., Frohme M., Furlan J.R.,  
 RA Ho P.L., Hohnselt J.D., Jungheira M.H.S., Gomes S.L., Gruber A.,  
 RA Klieger J.E., Kuramae E.E., Laigret F., Lambert M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsukuma A.Y.,  
 RA Marques D.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Panal A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;  
 "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 Nature 406:151-157(2000).  
 RT EMBL: AF001859; BAA94781.1;  
 DR INTERPRO: IPR002103;  
 DR PFAM: PF00296; Bac\_luciferase; 1  
 SQ SEQUENCE 321 AA; 35575 MW; F708E2A51F076E6 CRC64;

Query Match  
 Best Local Similarity 5.4%; Score 7; DB 2; Length 321;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 AERLAIR 40  
 Db 226 AERLAIR 232

RESULT 10  
 ID 091 PRELIMINARY; PRT; 341 AA.  
 AC 091MK8;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE HYPOTHETICAL PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC  
 clone: p0485D9.1";  
 RL Submitted (AFR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF001859; BAA94781.1;  
 DR PFAM: PF00092; vwa; 1.  
 DR PROSITE: PS50041; C-type\_LECTIN\_2; 1.  
 SQ SEQUENCE 341 AA; 39305 MW; D05403F7085726F6 CRC64;

OY 101 IYIYIKT 107  
 Db 308 IYIYIKT 314

RESULT 11  
 ID 085914 PRELIMINARY; PRT; 347 AA.  
 AC 085914;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DE HYPOTHETICAL 38.5 KDA PROTEIN.  
 OS Sphingomonas aromaticivorans.  
 OC Plasmid pWL1.  
 CC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
 OX Sphingomonas.  
 NCBI\_TaxID=48935;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-F199;  
 RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,  
 RA Sensen C.W., Gaasterland T., Saifer J.D., Fredrickson J.K.;  
 RT "Complete sequence of a 184 kb catabolic plasmid from *Sphingomonas*  
 RT aromaticivorans strain F199,"  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF079317; AAD03929.1;  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 347 AA; 38547 MW; A14E80670132F89F CRC64;

Query Match  
 Best Local Similarity 5.4%; Score 7; DB 2; Length 347;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 AOAERLA 38  
 Db 328 AOAERLA 334

RESULT 12  
 ID 023568 PRELIMINARY; PRT; 400 AA.  
 AC 023568; Q23575;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-MAY-1996 (TREMBlrel. 10, Last sequence update)  
 DE ZK673.9 PROTEIN.  
 GN ZK673.9  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderiinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wilkinson J.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z49132; CAA88991.1;  
 DR EMBL: Z48585; CAA88991.1; JOINED.  
 DR EMBL: Z49132; CAA88487.1;  
 DR EMBL: Z49132; CAA88487.1; JOINED.  
 DR INTERPRO: IPR001304;  
 DR INTERPRO: IPR002035;  
 DR PFAM: PF00092; vwa; 1.  
 DR PROSITE: PS50041; C-type\_LECTIN\_2; 1.  
 SQ SEQUENCE 400 AA; 44338 MW; 7C85723D679BA6B3 CRC64;

Query Match  
 Best Local Similarity 5.4%; Score 7; DB 5; Length 400;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPRKYP 8

Db 303 SFPKYP 309

RESULT 13

ID 060826 PRELIMINARY; PRT: 627 AA.  
AC 060826;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE JMI PROTEIN.  
GN JMI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PERIPHERAL BLOOD LEUKOCYTES;  
RA Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,  
RA Meindl A.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Blechschmidt K., Nyakatura G., Strom T.M., Drescher B., Menzel U.,  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF235097; AAF62519.1; -;  
DR EMBL: AF235097; AAF62519.1; -;  
SQ SEQUENCE 627 AA; 70755 MW; C3704BCD9EA0386 CRC64;

Query Match 5.4%; Score 7; DB 4; Length 627;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 119 GKIDRTP 125  
DB 530 GKIDRTP 536

RESULT 14  
Q9JIG7  
ID 09JIG7 PRELIMINARY; PRT: 627 AA.  
AC 09JIG7;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE XIMX40E PROTEIN.  
GN XIMX40E.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20313888; PubMed=10857745;  
RA Means G.D., Toy D.Y., Baum P.R., Derry J.M.J.;  
RT "A transcript map of a 2-Mb BAC contig in the proximal portion of the  
mouse X chromosome and regional mapping of the scurfy mutation.";  
RL Genomics 65:213-223(2000).  
DR EMBL: AF229637; AAF66951.1; -;  
SQ SEQUENCE 627 AA; 70843 MW; 3710732B4CD90A2E CRC64;

Query Match 5.4%; Score 7; DB 11; Length 627;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 119 GKIDRTP 125  
DB 530 GKIDRTP 536

RESULT 15  
Q9Y8A4  
ID 09Y8A4 PRELIMINARY; PRT: 636 AA.  
AC 09Y8A4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE 5-AMINOLEVULINIC ACID SYNTHASE.  
GN HEMA.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;  
OC anamorphic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AL560;  
RA Elrod S.L., Jones A., Cherry J.R.;  
RT "Cloning of 5-aminolevulinic synthase from Aspergillus oryzae and its  
use as a selectable marker.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF152374; AAD38391.1; -;  
DR INTERPRO: IPR001917; -;  
DR PFAM: PF00222; aminotran\_2; 1.  
DR PROSITE: PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
SQ SEQUENCE 636 AA; 68145 MW; F70BC2A073AOCDBA CRC64;

Query Match 5.4%; Score 7; DB 3; Length 636;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 PSSLRTL 14  
DB 20 PSSLRTL 26

Search completed: March 19, 2001, 14:56:34  
Job time: 66 sec

Mon Mar 19 15:04:00 2001

us-09-726-899-3.oligo.rspt

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2001, 14:55:23 ; Search time 16.08 seconds  
(without alignments)  
274.316 million cell updates/sec

Title: US-09-726-899-3

Perfect score: 129  
Sequence: 1 MSFPRKPSLSRLPTLPETLDP.....DRKEKLIQEGKIDRTFHLST 129

Scoring table:

GAPOP 60.0 , Gapext 60.0

Searched: 268485 seqs, 34193795 residues

Woblike: 0

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

A\_Geneseq\_36:\*

- 1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	129	19	W69225
2	95	73.6	113	20	W76629
3	7	5.4	449	19	W30559
4	7	5.4	636	19	W30557
5	7	5.4	636	19	W41509
6	7	5.4	636	19	W41498
7	7	5.4	928	21	W77292
8	7	5.4	928	21	W78844
9	7	5.4	3170	20	W39299
10	6	4.7	13	16	W21312
11	6	4.7	15	15	W45734
12	6	4.7	15	20	W95534
					Bj38 peptide fragm

13	6	4.7	16	21	Y70048
14	6	4.7	36	17	R95643
15	6	4.7	82	20	Y07954
16	6	4.7	83	20	Y34603
17	6	4.7	92	19	W69355
18	6	4.7	117	20	Y35533
19	6	4.7	161	13	R29647
20	6	4.7	161	15	R55575
21	6	4.7	161	19	W41300
22	6	4.7	161	20	Y30168
23	6	4.7	195	20	Y00177
24	6	4.7	229	17	W06639
25	6	4.7	229	16	W06791
26	6	4.7	260	20	W95531
27	6	4.7	285	21	Y70038
28	6	4.7	300	19	W69443
29	6	4.7	324	18	W24971
30	6	4.7	324	20	W27300
31	6	4.7	330	20	Y34527
32	6	4.7	333	20	W86333
33	6	4.7	339	16	R77317
34	6	4.7	344	20	Y34172
35	6	4.7	344	21	Y66696
36	6	4.7	345	20	Y34526
37	6	4.7	346	20	Y34525
38	6	4.7	357	20	Y34394
39	6	4.7	368	20	Y23342
40	6	4.7	371	20	Y29193
41	6	4.7	375	19	W61993
42	6	4.7	407	18	W24970
43	6	4.7	407	17	Y27299
44	6	4.7	425	17	W05510
45	6	4.7	440	17	R94600

#### ALIGNMENTS

RESULT 1	
W69225	W69225 standard. Protein: 129 AA.
XX	
AC	W69225;
XX	
DT	18-FEB-1999 (first entry)
XX	
DE	NADH dehydrogenase subunit NDS-2.
XX	
KW	NADH dehydrogenase subunit; sympathetic nervous system disease; myopathy;
KW	nicotinamide-adenine dinucleotide dehydrogenase; cancer; leukaemia;
XX	immune system disorder; neurodegenerative disease; therapy; NDS-2.
OS	Homo sapiens.
XX	
PN	W09831815-AA2.
XX	
PD	23-JUL-1998.
XX	
PF	17-DEC-1997; 97WO-US23970.
XX	
PR	17-JAN-1997; 97US-0785065.
XX	
PA	(INCY-) INCYTE PHARM INC.
XX	
PI	Bandman O, Goll SK, Hillman JL;
XX	
XX	WPI; 1998-414112/35.
DR	N-PSDB; V44787.
XX	
PT	Human nicotinamide-adenine di-nucleotide dehydrogenase subunits -
PT	useful for, e.g. diagnosis, treatment and prevention of cancer,
XX	myopathy, immune system disease and neuro-degeneration

PS Claim 19; Fig 2; 80pp; English.

CC This sequence represents the NMDH (reduced nicotinamide-adenine  
CC dinucleotide) dehydrogenase subunit, NDS-2, of the invention. Cells  
CC containing the DNA are used to produce the recombinant subunits.  
CC Antagonists of NDS-1 (typically antisense sequences or ribozymes) are  
CC used to treat or prevent cancer (leukemia and solid cancers) are  
CC system disorders (e.g. asthma, diabetes, rheumatoid arthritis, and immune  
CC osteoporosis and many others); NDS-2 and NDS-4 are used to treat myopathy  
CC their antagonists are used to treat cancer and lactic acidosis), while  
CC nervous system (e.g. hypertension, arrhythmia and migraine); NDS-3 is  
CC also used to treat myopathy and its antagonists to treat cancer and  
CC neurodegenerative disease (e.g. Alzheimer's, Huntington's and Parkinson's  
CC diseases, epilepsy and Down's syndrome). In all cases NDS or their  
CC antagonists may be expressed from gene therapy vectors. Ab may be used  
CC therapeutically as antagonist; as immunoassay reagent for diagnosis or  
CC monitoring such diseases; in competitive screening assays for agents that  
CC bind specifically to the subunits, and for affinity purification of the  
CC subunits from natural sources. The DNAs are useful as primers and probes  
CC for diagnosis and monitoring (including detecting predisposition to  
CC cancer); for gene mapping or identifying related sequences, while the  
CC subunits are also used to raise antibodies and to screen for specific  
CC binding agents.

Sequence 129 AA;

Query Match

Best Local Similarity 100.0%; Score 129; DB 19; Length 129;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPPKRSSLRPTLPDPAEYNISPEPRAQERLAIRAOQREYLLQNDPRKGLI 60  
DB 1 MSPPKRSSLRPTLPDPAEYNISPEPRAQERLAIRAOQREYLLQNDPRKGLI 60  
OY 61 ENPALLRMAYARTINVPNFRPTKNSLGMALCGRPILFIYIIRTERDKKELIOEGK 120  
DB 61 ENPALLRMAYARTINVPNFRPTKNSLGMALCGRPILFIYIIRTERDKKELIOEGK 120  
OY 121 LDRTFHLASY 129  
DB 121 LDRTFHLASY 129

RESULT 2

ID 176629 standard; Protein: 113 AA.  
AC 176629;

XX 10-APR-2000 (first entry)  
DE Human ovarian tumor EST fragment encoded protein 125.

XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;  
KW gene therapy; treatment.  
OS Homo sapiens.

XX DE19817557-A1.  
PN 21-OCT-1999.

XX 09-APR-1998; 98DE-1017557.  
PF 09-APR-1998; 98DE-1017557.

XX 09-APR-1998; 98DE-1017557.  
PR 09-APR-1998; 98DE-1017557.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX WPI: 1999-591920/51.  
DR

DR N-PSDB; 277502.

XX New nucleic acid sequences expressed in ovarian, and some other, cancer  
PT tissues, and derived polypeptides, for treatment of ovarian cancer and  
PT identification of therapeutic agents  
PS Claim 25; Page 295; 310pp; German.

CC This invention describes novel nucleic acid (cDNA) sequences (A) which  
CC have anticancer activity and are highly expressed in ovarian tumor  
CC tissue (and some also in testis and breast cancer tissue). The products  
CC of the invention can be used for gene therapy. (A) are used (i) for  
CC recombinant expression of polypeptides (B) and (ii) to isolate complete  
CC ovarian cancer; (ii) directly for treating this form of cancer (including  
CC expression from gene therapy vectors) and (iii) for generation of  
CC specific antibodies. (A) are identified by assembling ESTs (expressed  
CC sequence tags) from a particular tissue type before comparison of  
CC expression patterns. This allows a significantly longer fragment of the  
CC gene to be revealed, so should reduce the number of failures associated  
CC with the fact that ESTs from different libraries may represent different  
CC parts of the same unknown gene, distorting the estimated frequency of  
CC occurrence in a particular tissue. Y76505-Y76638 represent protein  
CC fragments encoded by the human ovarian tumor cDNA library derived EST  
CC fragments represented in 277450-277572.

Sequence 113 AA;

Query Match

Best Local Similarity 73.6%; Score 95; DB 20; Length 113;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 ERLAIRAOQREYLLQNDPRKGLIENPALLRMAYARTINVPNFRPTKNSLGMALCG 94  
DB 19 ERLAIRAOQREYLLQNDPRKGLIENPALLRMAYARTINVPNFRPTKNSLGMALCG 94  
OY 95 FGPLIFIIYIIRTERDKKELIOEGKLDRTFHLASY 129  
DB 79 FGPLIFIIYIIRTERDKKELIOEGKLDRTFHLASY 113

RESULT 3

ID W30559 standard; Protein: 449 AA.  
AC W30559;

XX 18-JAN-1999 (first entry)  
DE Aspergillus oryzae hema deletion allele-encoded protein.

XX 5-Aminoolevulinic acid synthase; hema gene; respiratory deficient;  
KW oxidative phosphorylation.  
OS Aspergillus oryzae strain A1560 (IFO 4177).

XX Synthetic.  
PN WO9841640-A1.

XX 24-SEP-1998.  
PD 17-MAR-1998; 98WO-US05156.

XX 17-MAR-1997; 97US-0819458.  
PR 17-MAR-1997; 97US-0819458.

XX (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (NOVO ) NOVO-NORDISK AS.

XX Cherry JR, Elrod SL, Jensen EB;  
PI WPI: 1998-521226/44.  
DR N-PSDB; V45426.



XX Production of polypeptide in respiratory-deficient cells transformed  
PT with construct - that complements the defect and encodes  
PT polypeptide, also production of defective cells by disrupting gene  
PT essential for oxidative phosphorylation  
XX  
PS Example 8; Fig 8A-B; 57pp; English.  
XX  
CC This polypeptide is encoded by a hemadelta::pyrC allele (see  
CC V45427) of the invention that includes a deletion in the hema gene  
CC (see V45424) encoding 5-aminolevulinic acid synthase (ALAS, see  
CC W30557). A. oryzae How425 cells transformed with the deletion  
CC allele demonstrated 5-aminolevulinic acid auxotrophy. Transformation  
CC with wild-type hema rescued the hema deletion phenotype. A claimed  
CC method of producing a polypeptide comprises: (a) introducing into a  
CC respiratory-deficient cell mutant (i) one or more first nucleic  
CC acid sequences (NAS) which complement the respiratory defect and  
CC (ii) a second NAS which encodes the polypeptide, (b) cultivating  
CC the cell aerobic conditions, and (c) isolating the polypeptide.  
CC The first NAS encodes a component of the electron transport chain,  
CC or an enzyme involved in the biosynthesis of ubiquinone, flavin or  
CC haem (e.g. ALAS). Also claimed is the respiratory-deficient mutant  
CC cell (preferably a haem deficient cell) and a method of producing  
CC such a cell. The method is used to select and maintain transformed  
CC cells, to ensure high level expression and genetic stability of  
CC transformants during culture. It is applicable to all industrial  
CC fermentation processes, requiring only that the cells need oxygen  
CC for growth.  
CC  
SQ Sequence 449 AA:

Query Match 5.4%; Score 7; DB 19; Length 449;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSLRRL 14  
| | | | | | |  
Db 20 psslrll 26

RESULT 4  
W30557  
ID W30557 standard; Protein; 636 AA.  
AC W30557;  
XX  
XX

DT 18-JAN-1999 (first entry)  
XX

DE Aspergillus oryzae 5-aminolevulinic acid synthase.  
XX

KM 5-Aminolevulinic acid synthase; hema gene; respiratory deficient;  
oxidative phosphorylation.  
XX

OS Aspergillus oryzae strain A1560 (IFO 4177).  
XX

FT Key Location/Qualifiers  
FT Peptide 1..35  
/note="putative mitochondrial leader"  
FT

PN W09841640-A1.  
XX

PD 24-SEP-1998.  
XX

PF 17-MAR-1998; 98WO-US05156.  
XX

PR 17-MAR-1997; 97US-0819458.  
XX

PA (NOVO) NOVO NORDISK BIOTECH INC.  
(NOVO) NOVO-NORDISK AS.  
XX

PI Cherry JR, Elrod SL, Jensen EB;  
XX

DR WPI; 1998-521226/44.  
DR N-PDB; V45424.  
XX  
XX Production of polypeptide in respiratory-deficient cells transformed  
PT with construct - that complements the defect and encodes  
PT polypeptide, also production of defective cells by disrupting gene  
PT essential for oxidative phosphorylation  
XX  
PS Example 5; Fig 3A-B; 57pp; English.  
XX

CC This is the amino acid sequence of 5-aminolevulinic acid synthase  
CC (ALAS), an enzyme of the haem biosynthetic pathway, that can be  
CC isolated from Aspergillus oryzae strain A1560. The sequence was  
CC deduced from an isolated genomic DNA (see V45424). A claimed  
CC method of producing a polypeptide comprises: (a) introducing into a  
CC respiratory-deficient cell mutant (i) one or more first nucleic  
CC acid sequences (NAS) which complement the respiratory defect and  
CC (ii) a second NAS which encodes the polypeptide, (b) cultivating  
CC the cell in a culture medium under aerobic conditions suitable for  
CC expression of the first and second NAS, and (c) isolating the  
CC polypeptide from the culture medium. The first NAS encodes a  
CC component of the electron transport chain, or an enzyme involved in  
CC the biosynthesis of ubiquinone, flavin or haem (e.g. ALAS). Also  
CC claimed is a respiratory-deficient mutant cell that is haem  
CC deficient (see V45426). The method is used to select and maintain  
CC transformed cells, to ensure high level expression and genetic  
CC stability of transformants during culture. It is applicable to all  
CC industrial fermentation processes, requiring only that the cells  
CC need oxygen for growth.  
CC  
SQ Sequence 636 AA:

Query Match 5.4%; Score 7; DB 19; Length 636;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSLRRL 14  
| | | | | | |  
Db 20 psslrll 26

RESULT 5  
W41509  
ID W41509 standard; Protein; 636 AA.  
AC W41509;  
XX  
XX

DT 22-JUN-1998 (first entry)  
XX

DE Aspergillus oryzae 5-aminolevulinic acid synthase.  
XX

KM 5-Aminolevulinic acid synthase; hema gene; herbicide; haemoprotein.  
XX

OS Aspergillus oryzae strain A1560 (IFO 4177).  
XX

FT Key Location/Qualifiers  
FT Peptide 1..35  
/label="Sig-peptide"  
/note="mutative mitochondrial leader"  
FT Peptide 11..15  
/note="haem regulatory motif"  
FT Peptide 50..54  
/note="haem regulatory motif"  
FT Binding-site 220..230  
/note="glycine loop involved in pyridoxal  
phosphate cofactor binding"  
FT

PN W09747736-A1.  
XX

PD 18-DEC-1997.  
XX

PF 09-JUN-1997; 97WO-US09928.  
XX

```

XX 10-JUN-1996; 96US-0019399.
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX Cherry JR, Elrod SL;
XX WPI: 1998-052303/05.
XX N-PSDB: V04102.
XX New 5-aminolevulinic acid synthase from Aspergillus oryzae - used
XX to produce 5-aminolevulinic acid herbicide and to increase
XX haemoprotein synthesis
XX Claim 1; Page 26-27; 62pp; English.
XX This protein comprises the 5-aminolevulinic acid synthase (ALA
XX synthase) of Aspergillus niger IFO 4177. Its amino acid sequence
XX was deduced from the isolate hema gene (see V04102) and shares
XX 91% identity with the Aspergillus nidulans ALA synthase. The
XX enzyme catalyses the reaction of glycine and succinyl-CoA to form
XX 5-aminolevulinic acid, which is useful as a herbicide. It is also
XX the rate-determining enzyme in biosynthesis of haem in liver cells
XX and differentiating erythrocytes, so overexpression in a cell can
XX be used to increase the yield of haemoproteins produced by the cell
XX in the absence of haem supplement. Host cells, especially
XX bacterial, fungal, filamentous fungal and yeast cells, transformed
XX with a vector that includes the hema gene are claimed, and can be
XX used in a claimed method for producing ALA synthase.
XX Sequence 636 AA:
SQ

```

Query Match 5.4%; Score 7; DB 19; Length 636;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 8 PSSRLTL 14
   |||||
DB 20 pssrltl 26

```

RESULT 6  
W41498  
ID W41498 standard; Protein: 636 AA.  
XX AC W41498;  
XX DT 1-JUN-1998 (first entry)  
XX DE 5-aminolevulinic acid synthase.  
XX KW 5-aminolevulinic acid synthase; porphobilinogen synthase;  
XX KM haemoprotein production; filamentous fungus; haeme biosynthetic enzyme.  
XX OS Aspergillus oryzae.  
XX PN W09747746-A1.  
XX PD 18-DEC-1997.  
XX PF 09-JUN-1997; 97WO-US10003.  
XX PR 17-MAR-1997; 97US-0041158.  
XX PR 10-JUN-1996; 96US-0662752.  
XX PA (NOVO ) NOVO NORDISK BIOTECH INC.  
XX PI Cherry JR, Elrod SL, Jones A;  
XX DR WPI: 1998-052312/05.  
XX DR N-PSDB: V19709.  
XX

```

PT Method for the production of haemo:protein in a filamentous fungal
PT cell - by introducing a haeme biosynthetic enzyme coding sequence
PT and control sequences into the cell
XX Claim 9; Page 45-46; 113pp; English.
XX This sequence is the 5-aminolevulinic acid synthase of Aspergillus
XX oryzae. The DNA can be used in the method of the invention for
XX producing a haemoprotein. The method comprises: (a) introducing
XX into a filamentous fungal cell: (i) one or more control sequences capable
XX of directing the expression of a haeme biosynthetic enzyme encoded by a
XX nucleic acid sequence endogenous to the filamentous fungal cell, where
XX one or more of the control sequences are operably linked to the nucleic
XX acid sequence; and/or (ii) one or more copies of one or more second
XX nucleic acid sequences encoding a haeme biosynthetic enzyme; (b)
XX cultivating the filamentous fungal cell in a nutrient medium suitable for
XX production of the haemoprotein and the haeme biosynthetic enzymes; and
XX (c) recovering the haemoprotein from the nutrient medium of the
XX filamentous fungal cell. The method is used to yield commercially
XX significant quantities of haemoprotein in filamentous fungal strains.
XX Sequence 636 AA:
SQ

```

Query Match 5.4%; Score 7; DB 19; Length 636;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 8 PSSRLTL 14
   |||||
DB 20 pssrltl 26

```

RESULT 7  
Y77292  
ID Y77292 standard; Protein: 928 AA.  
XX AC Y77292;  
XX DT 22-MAY-2000 (first entry)  
XX DE Streptomyces antibioticus oleandomycin PKS loading module.  
XX KW Polyketide; macrolide; biosynthesis; polyketide synthase; PKS;  
XX KM multienzyme complex; loading module; ketosynthase domain; KS; CLF domain;  
XX KW decarboxylation; acyl carrier protein domain; ACP; antihelmintic;  
XX insecticide; immunosuppressant; antifungal; antibacterial; oleandomycin.  
XX OS Streptomyces antibioticus.  
XX PN W0200000618-A2.  
XX PD 06-JAN-2000.  
XX PF 29-JUN-1999; 99WO-GB02044.  
XX PR 29-JUN-1998; 98GB-0014006.  
XX PA (BIOR-) BIOTICA TECHNOLOGY LTD.  
XX PI Leadlay PF, Staunton J, Cortes J, McArthur HAI;  
XX DR WPI: 2000-170919/15.  
XX PT Novel methods for preparing new variant polyketides, for use as  
XX PT antihelmintics, insecticides, immunosuppressants, antifungals or  
XX PT antibacterials -  
XX PS Claim 7; Fig 4A-C; 97pp; English.  
XX The invention relates to a novel system for producing polyketides  
XX particularly 12-, 14- and 16-membered ring macrocides from a desired  
XX starter unit. The biosynthesis of polyketides is initiated by a group  
XX

CC of chain-forming enzymes known as polyketide synthases (PKSs) which are  
 CC multienzyme complexes consisting of a set, or module, of enzymes  
 CC which catalyze polyketide chain extension. The system of the  
 CC invention comprises inserting loading modules into PKSs that do not  
 CC normally possess them, thereby controlling the starter units used. The  
 CC loading module may be adapted to load an optionally substituted malonyl  
 CC residue, which it then decarboxylates to provide an optionally  
 CC substituted acetyl residue for transfer to a chain extension module. The  
 CC loading module comprises a KS (ketosynthase)-type domain which effects  
 CC decarboxylation, and an acyl carrier protein domain (ACP). The KS-type  
 CC domain is preferably a Ksq domain, which possesses a glutamine rather  
 CC than a cysteine in the active site. Alternatively a Clp-type domain,  
 CC which also contains a glutamine at this site, may provide the  
 CC decarboxylating functionality. The methods of the invention are used to  
 CC produce polyketides, particularly 12-, 14- and 16-membered ring  
 CC macrolides. The system is used to produce macrolides with preferred  
 CC (acetate or propionate) starter units, or with unusual starter units,  
 CC which minimizes the formation of by-products containing a different  
 CC starter unit. The polyketides produced have use as potential  
 CC antihelmintics, insecticides, immunosuppressants, antifungals or  
 CC antibacterials. The present invention provides a system for producing an  
 CC polyketides which minimizes the formation of by-products containing an  
 CC undesired or different starter units, and also allows the incorporation  
 CC of unusual starter units. The system allows the identification of  
 CC polyketides which may have enhanced properties or possess novel  
 CC bioactivity. Sequences Y77289-Y77293 respectively represent the  
 CC loading domains of Streptomyces caelestis niddamycin PKS, S. ambofaciens  
 CC spiramycin (platenolide), S. cinnamonensis monensin PKS, S. antibioticus  
 CC oleandomycin PKS and S. fradiae tylosin PKS.

XX Sequence 928 AA:

Query Match 5.4%; Score 7; DB 21; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 RAQAERL 37  
 |||||  
 Db 487 raqaerl 493

RESULT 8  
 Y78844 Y78844 standard; Protein; 928 AA.

XX AC Y78844;

XX DE 9-MAY-2000 (first entry)

XX KSq-Atq loading didomains of oleandomycin PKS.

KW 14-member macrolide; PKS; antibiotic; polyketide synthase; production;  
 KM acetate starter unit; oleandomycin; Ksq-Atq loading domain.

XX Saccharopolyspora erythraea.

XX PN WO20000500-A2.

XX PD 06-JAN-2000.

XX PF 29-JUN-1999; 99WO-GB02042.

XX PR 29-JUN-1998; 98GB-0014006.

XX PA (BIOT-) BIOTICA TECHNOLOGY LTD.  
 (PRIZ ) PRIZER INC.

XX PI Leadlay PF, Staunton J, Cortes J, McArthur HAI;

XX DR WPI: 2000-170901/15.

XX New 14-member macrolides incorporating acetate starter unit, used as

PT antibiotics -  
 XX Disclosure; Fig 4; 78pp; English.

CC This sequence represents the Ksq-Atq loading didomain amino acid sequence  
 CC of the oleandomycin polyketide synthase (PKS). PKS is used in a system  
 CC for the production of the macrolides of the invention. The macrolides are  
 CC 14-member macrolides that incorporate an acetate starter unit so that it  
 CC has a 13-methyl substituent, provided that it is not noreerythromycin C,  
 CC 6-deoxy-15-noreerythromycin B or 6-deoxy-15-noreerythromycin D. The new  
 CC 14-member macrolides may be used as antibiotics. The macrolides are  
 CC produced by a process which minimizes the formation of by-products  
 CC containing different starter units. 13-Methyl erythromycins can be  
 CC produced at good expression levels and in substantial absence of  
 CC erythromycins with different starter units. Chemical modifications  
 CC previously only possible with 'natural' erythromycins can be performed.

SQ Sequence 928 AA:

Query Match 5.4%; Score 7; DB 21; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 RAQAERL 37  
 |||||  
 Db 487 raqaerl 493

RESULT 9  
 Y39299 Y39299 standard; Protein; 3170 AA.

XX AC Y39299;  
 XX 01-DEC-1999 (first entry)

XX SpnC a polyketide synthase.

KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal  
 KM microlides; arachnid; nematode; insect; polyketide; polyketide synthase;  
 KM PKS; extender module; initiator module; acyl transferase domain; AT;  
 KM acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;  
 KM dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;  
 KM insecticide.

XX Saccharopolyspora spinosa.

XX Key Location/Qualifiers

XX FH 1..423

XX FT /label= KS3

XX FT /note= "beta-ketosynthase domain: part of extender

XX FT module 3"

XX FT 531..850

XX FT /label= AT3

XX FT /note= "Acyl transferase domain: part of extender

XX FT module 3"

XX FT 1159..1337

XX FT /label= KR3

XX FT /note= "beta-ketoreductase domain: part of extender

XX FT module 3"

XX FT 1425..1506

XX FT /label= ACP3

XX FT /note= "Acyl carrier protein domain: part of extender

XX FT module 3"

XX FT 1529..1952

XX FT /label= KS4

XX FT /note= "beta-ketosynthase domain: part of extender

XX FT module 4"

XX FT 2066..2396

XX FT /label= AT4

XX FT /note= "Acyl transferase domain: part of extender

XX FT module 4"

FT Domain 2700..2880  
 FT /label= KR4  
 FT /note= "beta-ketoreductase domain: part of extender  
 FT module 4"  
 FT 2972..3053  
 FT /label= ACP4  
 FT /note= "Acyl carrier protein domain: part of extender  
 FT module 4"  
 PN MO9946387-A1.  
 PD 16-SEP-1999.  
 PE 16-FEB-1999; 99WO-US03212.  
 PR 09-MAR-1998; 98US-0036987.  
 PA (DOMC) DOM AGROSCIENCES LLC.  
 PI Raltz RH, Broughton MC, Crawford KP, Madduri K, Merlo DJ;  
 PI Readway PJ, Turner JR, Waldron C;  
 DR MPI; 1999-551414/46.  
 DR N-PDB; 221501.  
 XX New spinosyn biosynthetic genes from *Saccharopolyspora spinosa*, useful  
 PT for production of insecticidal spinosyn compounds  
 PS Claim 1; Page 103-113; 190pp; English.  
 XX This is the amino acid sequence of the product of the spnc gene. The  
 CC protein is involved in spinosyn biosynthesis. The spnc gene is one of 23  
 CC genes and open reading frames contained in an 80kb DNA sequence 221501.  
 CC Spinosyns are insecticidal microbicides which are useful for the control of  
 CC arachnids, nematodes and insects. Biosynthesis of spinosyns occurs via  
 CC stepwise condensation and modification of carboxylic acid precursors  
 CC generating a linear polyketide which is modified further. The DNA  
 CC sequence contains a central region of approximately 55kb which has  
 CC homology to the DNA encoding the polyketide synthases (PKS) of known  
 CC macrocyclic producers. The spinosyn PKS DNA region consists of 5 ORFs with  
 CC stop codons at the end of acyl carrier protein (ACP) domains. Together  
 CC the PKS polyketides (Y39297-Y39301) of which this sequence is one, form  
 CC a complex consisting of an initiator module, spnA, and several extender  
 CC modules spnB-spnE. Each extender module adds a specific acetyl Co-A  
 CC precursor to a growing polyketide chain, and modifies the beta-keto group  
 CC in a specific manner. A module in a PKS polyketide consists of several  
 CC domains with specific functions. The initiator module has an acyl  
 CC transferase (AT) domain, and an acyl carrier protein (ACP) domain. The  
 CC extender modules have the same domains plus a beta-ketosynthase (KS)  
 CC domain and optionally a beta-ketoreductase domain, a dehydratase (DH)  
 CC domain, and an enoyl reductase (ER) domain. The last extender module  
 CC terminates with a thioester domain. The products of the genes present in  
 CC the upstream region of the PKS genes have been assigned names spnF-spnS  
 CC Y39302-Y39315 and are responsible for different modifications in spinosyn  
 CC biosynthesis. There are also two ORFs ORF15 and ORF16 present  
 CC immediately upstream of spnS, producing polyketides Y39316-Y39317, and  
 CC two ORFs ORF1 and ORF2 present downstream of the PKS region producing  
 CC polyketides Y39318-Y39319. The genes are useful to improve yields of  
 CC spinosyns, and for creating new spinosyns e.g. by mutagenesis, or  
 CC interruption of steps in spinosyn biosynthesis. The modified spinosyns  
 CC may be a new insect control agent or serve as substrates for further  
 CC chemical modification and the creation of new semi-synthetic spinosyns.  
 CC The genes are also useful to isolate similar sequences from *S. spinosa* or  
 CC other species by hybridization.  
 XX Sequence 3170 AA;

Query Match 5.4%; Score 7; DB 20; Length 3170;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 31 RAOAERL 37

Db 491 raqer1 497  
 RESULT 10  
 ID W21312  
 ID W21312 standard; peptide; 13 AA.  
 AC W21312;  
 DT 29-JUL-1997 (first entry)  
 DE Glucagon precursor derived signal oligopeptide #17.  
 XX Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; HIV;  
 KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor;  
 KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
 KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
 KW gonadoliberin precursor; plasminogen activator inhibitor 2; protein;  
 KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
 KW apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
 KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
 KW treponema pallidum membrane protein; TmPA; islet amyloid polypeptide;  
 KW fibroblast MMP1; schistosoma elastase precursor; schistosoma;  
 KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.  
 OS Homo sapiens.  
 XX W09519568-A1.  
 PD 20-JUL-1995.  
 XX 12-JAN-1995; 95WO-US00575.  
 PF 14-JAN-1994; 94US-0182248.  
 PR (RATH/) RATH M.  
 PA Rath M;  
 PI MPI; 1995-263953/34.  
 DR Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
 PT regions of max. hydrophilicity, used in modulating communication  
 PT between protein(s)  
 PS Claim 5; Page 41; 88pp; English.  
 XX The sequences given in W21201-560 represent hydrophilic signal oligo-  
 CC peptides. These signal oligopeptides are localised on the surface  
 CC of the protein and are represented by the hydrophilicity maxima of  
 CC the protein. These peptides are enriched in charged amino acids  
 CC arranged with neutral spacer amino acids. The specific signal  
 CC character of these oligopeptides is determined by a characteristic  
 CC combination of conformation and charge within the signal sequence.  
 CC These oligopeptides may be used as vaccines in the treatment of  
 CC human disease, as competitive inhibitors to prevent or block the  
 CC metabolic action or interaction of a selected protein by blocking  
 CC its specific signal sequences, or as therapeutic agents to function  
 CC as feedback regulators to reduce synthesis rate of a selected protein.  
 CC These peptides may be modified by omitting one or more amino acids at  
 CC the N- and/or C-terminal, by substituting one or more amino acids  
 CC without consideration of charge and polarity, by substituting one or  
 CC more amino acids with amino acid residues with similar charge and/or  
 CC polarity, by omitting one or more amino acids or a combination of these.  
 XX Sequence 13 AA;

Query Match 4.7%; Score 6; DB 16; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 ETRRAQ 33  
 Db 7 etirraq 12

## RESULT 11

ID R45734 standard; peptide; 15 AA.

XX R45734;

DT 23-JUL-1994 (first entry)

XX Interleukin-6 mutein (residues 171-185).

KW IL-6: carboxy terminal mutants; muteins; proliferation;  
 KM differentiation; immunotherapeutic; antiinflammatory;  
 KW thrombocytopenia; chemotherapy; bone marrow transplant.

XX Synthetic.

XX R45734; 09402512-A.

XX 03-FEB-1994.

XX 23-JUL-1993; 93WO-US06928.

XX 23-JUL-1992; 92US-0918181.

XX (UNNC-) UNIV NORTH CAROLINA.

XX Fowlkes D;

XX WPI: 1994-048796/06.

PT New carboxy terminal interleukin-6 muteins - having amino acid  
 PT substns. at position 171 or 175, for use in immunotherapeutic or  
 PT anti-inflammation compns.

PS Disclosure: Fig 6; 79pp; English.

CC The sequence shows a carboxy mutein of interleukin 6 from residues  
 CC 171-185. It has been found that mutants of IL-6 having amino  
 CC acid substns. at amino acid 171 or 175 have increased activity over  
 CC the wild type sequence. The IL-6 muteins are useful in  
 CC proliferation of B cells, T cells, megakaryocytes and multi-  
 CC potential haematopoietic progenitor cells and they also induce  
 CC various acute phase proteins in liver cells. They are useful in  
 CC immunotherapeutic and antiinflammation compns. They can also be  
 CC used for the treatment of patients suffering from thrombocytopenia  
 CC or undergoing chemotherapy or bone marrow transplant.  
 CC See also R45717-39.

CC Sequence 15 AA;

Query Match 4.7%; Score 6; DB 15; Length 15;

Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 SSLRLT 14

Db 7 slrlt 12

## RESULT 12

ID W95534 standard; peptide; 16 AA.

XX W95534;

XX 24-MAR-1999 (first entry)

DE B338 peptide fragment deduced from DNA sequence analysis.  
 XX Carbohydrate-binding protein; B338; chromosomal DNA; soya root;  
 KM modulation; nitrogen fixation; soya bean.  
 XX

OS Bradyrhizobium japonicum.

XX US5863728-A.

XX 26-JAN-1999.

PF 05-JUL-1996; 96US-0675921.

XX 05-JUL-1996; 96US-0675921.

PA (UNMS ) UNIV MICHIGAN STATE.

PI Ho JS, Loh JT, Schindler MS, Wang JL;

XX WPI: 1999-131297/11.

PT DNA encoding lectin B338 - derived from Bradyrhizobium japonicum

PS Claim 15; Fig 1B; 26pp; English.

CC The invention relates to a DNA encoding a carbohydrate-binding protein  
 CC (B338), where the DNA is present in an EcoRI-BamHI segment of  
 CC Bradyrhizobium japonicum chromosomal DNA and is recognised by a probe  
 CC selected from a 794 base pair sequence or its complement (X00921 or  
 CC X00920 respectively). Bradyrhizobium japonicum strains transformed with  
 CC the DNA are more competitive in occupying soya root nodulation sites than  
 CC wild-type strains leading to an increase nitrogen fixation and soya bean  
 CC yields. The present sequence represents a partial amino acid fragment  
 CC of B338 deduced from DNA sequence analysis.

XX Sequence 16 AA;

Query Match 4.7%; Score 6; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 LAIRAQ 42

Db 3 lairraq 8

## RESULT 13

ID Y70048 standard; protein; 16 AA.

XX Y70048;

DT 05-JUN-2000 (first entry)

DE A. halophila SDMT peptide-6.

KW Sarcosine-dimethylglycine methyltransferase; SDMT; betaine operon;  
 KM dimethylglycine content; abiotic stress; tolerance; salt; freezing;  
 KW cold; drought stress; pathogenesis-related protein; animal feed;  
 KW pH tolerance; environmental stress; inclusion body formation.

XX Actinopolyspora halophila.

XX WO200011142-A2.

XX 02-MAR-2000.

PF 18-AUG-1999; 99WO-EP06037.

PR 20-AUG-1998; 98US-0137434.

XX (CULT-) CULTOR CORP.

XX Relinkainen T, Nyssöelae A, Kerovu J;  
 XX WPI: 2000-224686/19.  
 DR  
 XX  
 XX  
 PT New methyltransferases, useful for e.g. producing transgenic plants  
 PT with increased stress tolerance, pathogen resistance or nutritional  
 PT value as animal feed or for improving the viability of microorganisms  
 PT in the food industry -  
 PS  
 XX Example 6: Page 45, 176pp: English.  
 XX  
 CC The present sequence is Actinopolyspora halophila sarcosine  
 CC -dimethylglycine methyltransferase (SDMT) peptide used to make PCR  
 CC primers for isolation of A. halophila SDMT gene. This sequence was  
 CC determined by using Perkin Elmer/Applied Biosystems Procise 494A protein  
 CC sequencing system. Expression vectors comprising the coding region from  
 CC betaine operon can be used to increase intracellular betaine and  
 CC dimethylglycine content. The polynucleotide can be used to increase  
 CC salt, freezing or cold tolerance, increase resistance to drought stress,  
 CC pathogens or induce pathogenesis-related proteins in plants. Transgenic  
 CC organisms can be used as an animal feed ingredient. The polynucleotide  
 CC can also be used to enhance pH tolerance and improve viability of  
 CC organisms when subjected to environmental stress. This can decrease  
 CC inclusion body formation when used in conjunction with polynucleotides  
 CC encoding a heterologous protein.  
 CC  
 XX Sequence 16 AA:  
 SQ  
 Query Match 4.7%; Score 6; DB 21; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 25 ISPEPR 30  
 |||||  
 Db 3 ispepr 8

RESULT 14  
 R95643  
 ID R95643 standard; peptide; 36 AA.  
 XX  
 AC R95643;  
 XX  
 DT 18-DEC-1996 (first entry)  
 XX  
 DE Oxyntomodulin variant.  
 XX  
 KW Oxyntomodulin; intestine; eel; A. japonica; cardiosstimulant;  
 KW insulin release accelerator; heart disease; diabetes.  
 XX  
 OS Anguilla japonica.  
 XX  
 PN MO9614336-A1.  
 XX  
 PD 17-MAY-1996.  
 XX  
 PF 07-NOV-1995; 95WO-JP02269.  
 XX  
 PR 07-NOV-1994; 94JP-0272069.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Ando M, Moriyama T, Yamasaki M, Yano K;  
 XX  
 DR WPI: 1996-251715/25.  
 XX  
 PT New oxyntomodulin peptide from Anguilla japonica - is a  
 PT cardio-stimulant and insulin release accelerator, useful for  
 PT treatment of heart disease and diabetes  
 XX  
 PS Claim 1: Page 13; 20pp; Japanese.

XX This sequence represents a variant of oxyntomodulin. It is  
 CC extracted from the intestines of the eel Anguilla japonica. This  
 CC peptide is a cardiosstimulant and an insulin release accelerator. It  
 CC is useful in the treatment of heart disease and diabetes.  
 XX  
 XX Sequence 36 AA:  
 SQ  
 Query Match 4.7%; Score 6; DB 17; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 28 ETRRAQ 33  
 |||||  
 Db 15 etrraq 20

RESULT 15  
 Y07954  
 ID Y07954 standard; Protein; 82 AA.  
 XX  
 AC Y07954;  
 XX  
 DT 06-JUL-1999 (first entry)  
 XX  
 DE Human secreted protein fragment #3 encoded from gene 1.  
 XX  
 KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;  
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;  
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;  
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;  
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;  
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;  
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;  
 KW arthritis; malignancy; digestive; endocrine; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9918208-A1.  
 XX  
 PD 15-APR-1999.  
 XX  
 PF 01-OCT-1998; 98WO-US20775.  
 XX  
 PR 02-OCT-1997; 97US-0060884.  
 XX  
 PR 02-OCT-1997; 97US-0060833.  
 XX  
 PR 02-OCT-1997; 97US-0060836.  
 XX  
 PR 02-OCT-1997; 97US-0060837.  
 XX  
 PR 02-OCT-1997; 97US-0060838.  
 XX  
 PR 02-OCT-1997; 97US-0060839.  
 XX  
 PR 02-OCT-1997; 97US-0060843.  
 XX  
 PR 02-OCT-1997; 97US-0060862.  
 XX  
 PR 02-OCT-1997; 97US-0060866.  
 XX  
 PR 02-OCT-1997; 97US-0060874.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;  
 XX  
 PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;  
 XX  
 PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;  
 XX  
 DR WPI: 1999-264022/22.  
 XX  
 PT New isolated human genes and the secreted polypeptides they encode  
 XX  
 PS Disclosure: Page 325; 368pp: English.  
 XX  
 CC This invention describes novel isolated human genes and the secreted  
 CC proteins they encode. The products of the invention are useful for  
 CC preventing, treating or ameliorating medical conditions, e.g. by protein  
 CC or gene therapy. Also pathological conditions can be diagnosed by  
 CC determining the amount of the new polypeptides in a sample or by







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2001, 14:55:28 ; Search time 12.64 Seconds  
(without alignments)  
183.264 Million cell updates/sec

Title: US-09-726-899-3

Perfect score: 129

Sequence: 1 MSFPKXKPSLRLPETLDP.....DRKEKLIQEGKLDRTFHLISY 129

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 174772 seqs, 17957048 residues

Wc size: 0  
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129	100.0	129	2 US-08-785-065-3	Sequence 3, Appl
2	17	13.2	129	2 US-08-785-065-10	Sequence 10, Appl
3	7	5.4	449	2 US-08-819-458A-16	Sequence 16, Appl
4	7	5.4	636	2 US-08-871-266B-2	Sequence 2, Appl
5	7	5.4	636	2 US-08-819-458A-2	Sequence 2, Appl
6	7	5.4	636	2 US-09-018-864A-2	Sequence 2, Appl
7	7	5.4	636	2 US-08-871-267B-2	Sequence 2, Appl
8	7	5.4	3170	3 US-09-036-987A-4	Sequence 4, Appl
9	6	4.7	15	1 US-07-918-181A-30	Sequence 30, Appl
10	6	4.7	15	1 US-08-231-575-30	Sequence 30, Appl
11	6	4.7	15	4 PCT-US93-06928-30	Sequence 30, Appl
12	6	4.7	16	2 US-08-675-921B-6	Sequence 6, Appl
13	6	4.7	36	2 US-08-835-528-1	Sequence 1, Appl
14	6	4.7	92	1 US-08-446-038B-21	Sequence 21, Appl
15	6	4.7	92	1 US-08-446-010B-21	Sequence 21, Appl
16	6	4.7	92	2 US-08-479-078-18	Sequence 18, Appl
17	6	4.7	92	2 US-08-805-445-21	Sequence 21, Appl
18	6	4.7	92	2 US-08-064-067D-21	Sequence 21, Appl
19	6	4.7	92	2 US-09-066-208-23	Sequence 23, Appl
20	6	4.7	94	1 US-08-167-035-23	Sequence 23, Appl
21	6	4.7	94	1 US-08-208-887A-23	Sequence 23, Appl
22	6	4.7	94	1 US-08-539-005-23	Sequence 23, Appl
23	6	4.7	144	1 US-07-956-700B-89	Sequence 89, Appl
24	6	4.7	144	1 US-08-476-537-89	Sequence 89, Appl
25	6	4.7	144	1 US-08-485-607-89	Sequence 89, Appl
26	6	4.7	144	2 US-08-475-879-89	Sequence 89, Appl
27	6	4.7	161	1 US-07-991-867B-5	Sequence 5, Appl
28	6	4.7	161	1 US-08-107-755A-5	Sequence 5, Appl

29	6	4.7	161	2 US-08-544-332-5	Sequence 5, Appl
30	6	4.7	229	4 PCT-US96-03916-13	Sequence 13, Appl
31	6	4.7	229	4 PCT-US96-03916-62	Sequence 62, Appl
32	6	4.7	260	2 US-08-675-921B-9	Sequence 9, Appl
33	6	4.7	344	2 US-09-055-097-3	Sequence 3, Appl
34	6	4.7	425	1 US-08-414-826A-15	Sequence 15, Appl
35	6	4.7	425	2 US-08-926-922-15	Sequence 15, Appl
36	6	4.7	425	3 US-09-253-682-15	Sequence 15, Appl
37	6	4.7	515	2 US-08-705-660-46	Sequence 46, Appl
38	6	4.7	515	3 US-08-989-045-46	Sequence 46, Appl
39	6	4.7	578	1 US-08-766-014-4	Sequence 4, Appl
40	6	4.7	608	1 US-08-766-014-3	Sequence 3, Appl
41	6	4.7	617	1 US-08-279-700-2	Sequence 2, Appl
42	6	4.7	617	1 US-08-279-700-4	Sequence 4, Appl
43	6	4.7	617	1 US-08-279-700-6	Sequence 6, Appl
44	6	4.7	617	1 US-08-279-700-8	Sequence 8, Appl
45	6	4.7	617	1 US-08-279-700-10	Sequence 10, Appl

#### ALIGNMENTS

RESULT 1  
US-08-785-065-3  
Sequence 3, Application US/08785065  
Patent No. 5814451  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K  
TITLE OF INVENTION: NOVEL SUBUNITS OF NADH DEHYDROGENASE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,065  
FILING DATE: Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0187 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Consensus  
CLONE: Consensus  
US-08-785-065-3

Query Match 100.0%; Score 129; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 6, 3e-119;

Matches 129: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFPRKRSSTLRTLPETLDPDAEYNISPTERRAQERLAIROKREYLYQNDPRRGLI 60  
DB 1 MSFPRKRSSTLRTLPETLDPDAEYNISPTERRAQERLAIROKREYLYQNDPRRGLI 60  
QY 61 ENPALLRNAVARTINVYNFRTPKNSLMGALCGPLIFITYIITKTRDRKREKLIQSGK 120  
DB 61 ENPALLRNAVARTINVYNFRTPKNSLMGALCGPLIFITYIITKTRDRKREKLIQSGK 120  
QY 121 LDRTFHLSTY 129  
DB 121 LDRTFHLSTY 129

## RESULT 2

US-08-785-065-10  
Sequence 10, Application US/08785065  
Patent No. 5814451  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL SUBUNITS OF NADH DEHYDROGENASE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,065  
FILING DATE: Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0187 US  
ELECTRONIC COMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 114  
US-08-785-065-10

Query Match 13.2%; Score 17; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 1,6e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 RDRKREKLIQSGKLDRTF 125  
DB 109 RDRKREKLIQSGKLDRTF 125

## RESULT 3

US-08-819-458A-16  
Sequence 16, Application US/08819458A  
Patent No. 5891669  
GENERAL INFORMATION:  
APPLICANT: Jensen, Ejner B.  
APPLICANT: Cherry, Joel  
APPLICANT: Elrod, Susan L.  
TITLE OF INVENTION: Methods For Producing Polypeptides  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 58916690 No. 5891669disk of No. 5891669th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/819,458A  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5215, 000-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-819-458A-16

Query Match 5.4%; Score 7; DB 2; Length 449;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSLRTL 14  
DB 20 PSSLRTL 26

## RESULT 4

US-08-871-266B-2  
Sequence 2, Application US/08871266B  
Patent No. 5871991  
GENERAL INFORMATION:  
APPLICANT: Elrod, Susan L.  
APPLICANT: Cherry, Joel R.  
TITLE OF INVENTION: Aspergillus Oryzae 5-Aminolevulinic  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 58719910 No. 5871991disk of No. 5871991th America, Inc.  
STREET: 405 Lexington Avenue - 64th Fl.  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

QY 109 RDRKREKLIQSGKLDRTF 125  
DB 109 RDRKREKLIQSGKLDRTF 125

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,266B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-878-9652  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: Internal  
US-08-871-266B-2

Query Match 5.4%; Score 7; DB 2; Length 636;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSRLTL 14  
|||||  
Db 20 PSSRLTL 26

RESULT 5  
US-08-819-458A-2  
Sequence 2, Application US/08819458A  
Patent No. 5891669  
GENERAL INFORMATION:  
APPLICANT: Jensen, Ejner B.  
APPLICANT: Cherry, Joel L.  
APPLICANT: Elrod, Susan L.  
TITLE OF INVENTION: Methods for Producing Polypeptides  
TITLE OF INVENTION: In Respiratory-Deficient Cells  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58916690 No. 5891669disk of No. 5891669th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10174  
TELEFAX: 212-878-9655  
TELEX:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/819,458A  
FILING DATE: 17-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5215.000-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-819-458A-2

Query Match 5.4%; Score 7; DB 2; Length 636;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSRLTL 14  
|||||  
Db 20 PSSRLTL 26

RESULT 6  
US-09-018-864A-2  
Sequence 2, Application US/09018864A  
Patent No. 5958747  
GENERAL INFORMATION:  
APPLICANT: Elrod, Susan L.  
APPLICANT: Cherry, Joel R.  
TITLE OF INVENTION: Aspergillus Oryzae 5-Aminolevulinic  
TITLE OF INVENTION: Acid Synthases And Nucleic Acids Encoding Same  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59587470 No. 5958747disk of No. 5958747th America, Inc.  
STREET: 405 Lexington Avenue - 64th Fl.  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
TELEFAX: 212-878-9655  
TELEX:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,864A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/871,266  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-878-9652  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 636 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: Internal  
US-09-018-864A-2

Query Match 5.4%; Score 7; DB 2; Length 636;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PSSRLTL 14  
|||||  
Db 20 PSSRLTL 26

RESULT 7

US-08-871-267B-2  
; Sequence 2, Application US/08871267B  
; Patent No. 6100057  
; GENERAL INFORMATION:  
; APPLICANT: Elrod, Susan L.  
; APPLICANT: Cherry, Joel R.  
; APPLICANT: Jones, Audrey  
; TITLE OF INVENTION: A Method for Increasing Hemoprotein  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 61000570 No. 6100057disk Of No. 6100057th America, Inc.  
; STREET: 405 Lexington Avenue - 64th Fl.  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/871,267B  
; FILING DATE: 9-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kozek, Carol E.  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4771.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-878-9652  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 636 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: Internal  
; US-08-871-267B-2

Query Match 5.4%; Score 7; DB 3; Length 636;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 8 PSSLRTL 14  
DB 20 PSSLRTL 26

RESULT 8  
US-09-036-987A-4  
; Sequence 4, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patil J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road

CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R.  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3170 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-036-987A-4

Query Match 5.4%; Score 7; DB 3; Length 3170;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 RQAQERL 37  
DB 491 RQAQERL 497

RESULT 9  
US-07-918-181A-30  
; Sequence 30, Application US/07918181A  
; Patent No. 5338833  
; GENERAL INFORMATION:  
; APPLICANT: Fowlkes, Dana M.  
; TITLE OF INVENTION: C-Terminal IL-6 Muteins  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Imclone Systems Incorporated  
; STREET: 180 Varick Street  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10014  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/918,181A  
; FILING DATE: 23-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sheets, Eric J.  
; REGISTRATION NUMBER: 30,326  
; REFERENCE/DOCKET NUMBER: FOW-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-645-1405  
; TELEFAX: 212-645-2054  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
US-07-918-181A-30

Query Match 4.7%; Score 6; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 SSLRTL 14  
111111  
Db 7 SSLRTL 12

RESULT 10  
US-08-231-575-30  
Sequence 30, Application US/08231575  
Patent No. 5565336

GENERAL INFORMATION:  
APPLICANT: Fowles, Dana M.  
TITLE OF INVENTION: C-Terminal IL-6 Muteins  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Imclone Systems Incorporated  
STREET: 180 Varlick Street  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10014

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,575  
FILING DATE: 22-APR-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/918,181  
FILING DATE: 23-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheets, Eric J.  
REGISTRATION NUMBER: 30,326  
REFERENCE/DOCKET NUMBER: FOW-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-645-1405  
TELEFAX: 212-645-2054

INFORMATION FOR SEQ. ID NO: 30:

SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
US-08-231-575-30

Query Match 4.7%; Score 6; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 SSLRTL 14  
111111  
Db 7 SSLRTL 12

RESULT 11  
PCT-US93-06928-30  
Sequence 30, Application PC/TUS9306928

GENERAL INFORMATION:  
APPLICANT: Fowles, Dana M.  
TITLE OF INVENTION: Carboxy Terminal IL-6 Muteins  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Imclone Systems Incorporated  
STREET: 180 Varlick Street  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 11014

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06928  
FILING DATE: 19930723

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/918,181  
FILING DATE: 23-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Felt, Irving N.  
REGISTRATION NUMBER: 28,601  
REFERENCE/DOCKET NUMBER: FOW-2-T  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-645-1405  
TELEFAX: 212-645-2054

INFORMATION FOR SEQ. ID NO: 30:

SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
PCT-US93-06928-30

Query Match 4.7%; Score 6; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 SSLRTL 14  
111111  
Db 7 SSLRTL 12

RESULT 12

US-08-675-921B-6  
Sequence 6, Application US/08675921B  
Patent No. 5863728

GENERAL INFORMATION:  
APPLICANT: John Su-Cheong Ho, John T. Loh, Melvin  
APPLICANT: Schindler and John L. Wang  
TITLE OF INVENTION: DNA Encoding Carbohydrate  
TITLE OF INVENTION: Binding Protein and Biological  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ian C. McLeod  
STREET: 2190 Commons Parkway  
CITY: Okemos  
STATE: Michigan  
COUNTRY: USA

ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch, 360  
MEDIUM TYPE: kb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS (version 3.3)  
SOFTWARE: wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,921B  
FILING DATE: 06/05/96  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-265  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: No. 5863728E  
INFORMATION FOR SEQ. ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE:  
DESCRIPTION: Peptide  
HYPOTHETICAL: Yes  
ANTI-SENSE: NO  
FRAGMENT TYPE: N/A  
ORIGINAL SOURCE:  
ORGANISM: N/A  
STRAIN: N/A  
INDIVIDUAL ISOLATE: N/A  
DEVELOPMENTAL STAGE: N/A  
HAPLOTYPE: N/A  
TISSUE TYPE: N/A  
CELL TYPE: N/A  
CELL LINE: N/A  
ORGANELLE: N/A  
IMMEDIATE SOURCE:  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: deduced partial amino acid  
NAME/KEY: sequence of BU38 DNA  
LOCATION: N/A  
IDENTIFICATION METHOD: sequencing  
OTHER INFORMATION: N/A  
PUBLICATION INFORMATION: N/A  
US-08-675-921B-6

Query Match 4.7%; Score 6; DB 2; Length 16;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 LAIRAO 42  
DB 3 LAIRAO 8

RESULT 13  
US-08-836-528-1  
Sequence 1, Application US/08836528  
Patent No. 5858975  
GENERAL INFORMATION:  
APPLICANT: Yano, Keiichi  
APPLICANT: Yamasaki, Motoo  
APPLICANT: Moriyama, Takahiro  
APPLICANT: Ando, Masaaki

TITLE OF INVENTION: NOVEL OXYNTOMODULIN  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO  
STREET: 277 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10172-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,528  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 272069/94  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Perry, Lawrence S.  
REGISTRATION NUMBER: 31865  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-2400  
TELEFAX: 212-758-2982  
TELEX: 236262  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-528-1

Query Match 4.7%; Score 6; DB 2; Length 36;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 ETRRAQ 33  
DB 15 ETRRAQ 20

RESULT 14  
US-08-446-038B-21  
Sequence 21, Application US/08446038B  
Patent No. 5658791  
GENERAL INFORMATION:  
APPLICANT: Wilks, Andrew F.; Ziemlecki, Andrew;  
APPLICANT: Harpur, Ailsa  
NUMBER OF INVENTION: No. 5658791e1 Protein Tyrosine Kinase  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,038B  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/064,067  
FILING DATE: 30-Jun-1993  
APPLICATION NUMBER: PCT/US91/08889  
FILING DATE: 26-No. 5658791-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: Australian PK3594/90  
FILING DATE: 28-No. 5658791-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: Australian 88229/91  
FILING DATE: 27-No. 5658791-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5658791man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5244  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-688-9200  
TELEFAX: 212-838-3884  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-446-038B-21

Query Match 4.7%; Score 6; DB 1; Length 92;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 GKLDRT 124  
|||||  
DB 4 GKLDRT 9

RESULT 15  
US-08-446-010B-21  
Sequence 21, Application US/08446010B  
Patent No. 5716818  
GENERAL INFORMATION:  
APPLICANT: Wilks, Andrew F.; Ziemiecki, Andrew;  
APPLICANT: Harpur, Ailsa  
TITLE OF INVENTION: No. 5716818el Protein Tyrosine Kinase  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felife & Lynch  
STREET: 805 Third Avenue  
City: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette, 3.5 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,010B  
FILING DATE: 19-May-1995  
CLASSIFICATION: 433  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/446,038  
FILING DATE: 19-May-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/064,067  
FILING DATE: 30-Jun-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08889  
FILING DATE: 26-No. 5716818-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: Australian PK3594/90  
FILING DATE: 28-No. 5716818-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: Australian 88229/91

FILING DATE: 27-No. 5716818-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Baer, Madeline F.  
REGISTRATION NUMBER: 36,437  
REFERENCE/DOCKET NUMBER: LUD 5244.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-688-9200  
TELEFAX: 212-838-3884  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-446-010B-21

Query Match 4.7%; Score 6; DB 1; Length 92;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 GKLDRT 124  
|||||  
DB 4 GKLDRT 9

Search completed: March 19, 2001, 14:56:48  
Job time: 80 sec

Mon Mar 19 15:03:57 2001

us-09-726-899-3.oligo.ra1

---